Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Parameter and Uncertainty Estimation in Groundwater Modelling

Jensen, Jacob Birk

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Jensen, J. B. (2003). *Parameter and Uncertainty Estimation in Groundwater Modelling*. The Hydraulics and Coastal Engineering Group, Dept. of Civil Engineering, Aalborg University. Series paper No. 23

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: May 07, 2023

PARAMETER AND UNCERTAINTY ESTIMATION IN GROUNDWATER MODELLING

Jacob Birk Jensen

Hydraulics & Coastal Engineering Laboratory
Department of Civil Engineering
AALBORG UNIVERSITY

Hydraulics & Coastal Engineering Laboratory
Department of Civil Engineering
Aalborg University
Sohngaardsholmsvej 57
DK–9000 Aalborg, Denmark

ISSN 0909–4296
SERIES PAPER No. 23

Parameter and Uncertainty Estimation in Groundwater Modelling

by

Jacob Birk Jensen

December 2002

Published 2003 by
Hydraulics & Coastal Engineering Laboratory
Aalborg University

Printed in Denmark by
Aalborg University

ISSN 0909–4296
SERIES PAPER No. 23

Preface

The present thesis, *Parameter and Uncertainty Estimation in Groundwater Modelling*, is submitted as one of the requirements for obtaining the degree of Ph.D. in accordance with Order No. 989 of 11 December 1992 issued by the Danish Ministry of Education.

The study was supported by Department of Civil Engineering, Aalborg University and was carried out in the period from September 1998 to December 2002 at Aalborg University (AAU) under the supervision of Associate Professor Kjeld Schaarup-Jensen and Associate Professor Michael Brorsen, Department of Civil Engineering, Aalborg University.

The author wishes to thank Kjeld Schaarup-Jensen and Michael Brorsen for good and thorough supervision and everybody in the department for making it a very nice place to work.

Special thanks to my wife Karen and my children Thea and Karoline for their patience and support throughout the study.

Aalborg, December 2002.

Jacob Birk Jensen.

Table of Contents

Preface i

Contents iii

List of Symbols vii

Summary xiii

Summary in Danish xv

1 Introduction 1

1.1 Application of groundwater models 2

1.1.1 Water quantity modelling 3

1.1.2 Water quality modelling 3

1.2 Calibration and uncertainty estimation 4

1.3 Objectives 9

1.4 Outline of the thesis 11

2 Groundwater modelling 13

2.1 Numerical discretisation 14

2.2 Geological representation 15

2.3 Implementation of rivers 15

2.4 Governing equations 17

2.4.1 Groundwater flow 17

2.4.2 Surface flow 17

2.5 Numerical formulation 19

2.5.1 Potential flow terms 20

2.5.2 Sink/source terms 22

2.5.3 Storage terms 22

2.5.4 Vertical flow under de-watered conditions 23

2.5.5 Initial and boundary conditions 23

2.5.6 Solving the finite difference equations 24
2.6 Transport modelling 24
2.6.1 Governing equation 24
2.6.2 Particle tracking method 25
2.6.3 Verification of flow and transport models 29

3 Classical calibration procedures **31**
3.1 Parameterisation 32
3.1.1 The continuum approach 32
3.1.2 Parameter scale 33
3.2 Observation data 35
3.2.1 Head data 35
3.2.2 River discharges 39
3.2.3 Concentration data 40
3.2.4 Tracer data (natural and artificial) 41
3.2.5 Subjective observations 41
3.2.6 Weighting of observations 41
3.3 Ill-posedness 42
3.4 Non-linearity and discontinuity 43
3.5 Least square method 44
3.6 Maximum likelihood method 46
3.7 Single versus multi-objective parameter estimation 46
3.8 Parameter statistics 48
3.9 State variable statistics 49
3.10 Solving the regression problem 50
3.10.1 Gauss-Newton minimisation 50
3.10.2 Other methods 51

4 The Generalized Likelihood Uncertainty Estimation methodology **53**
4.1 Likelihood measures 55
4.1.1 Traditional statistical likelihood measures 56
4.1.2 Traditional GLUE likelihood measures 57
4.1.3 Fuzzy likelihood measures 58
4.1.4 Inference functions 59
4.2 Designing the likelihood measure 62
4.3 Bayesian updating of prior parameter distributions 63
4.4 An example 64
4.5 Generation of random parameter sets 68

4.6 Concluding remarks 68

5 Case A: Capture zone modelling in a synthetic setup 71
5.1 Conceptual models 72
5.2 Reference model 73
5.3 Quantification of model error 74
5.4 Monte Carlo simulations 76
5.5 Observation data 78
5.6 GLUE analyses 79
5.6.1 Conditioning based on head data 79
5.6.2 Conditioning based on river discharge data 80
5.6.3 Conditioning based on head data and river discharge data 81
5.6.4 Influence of point likelihood function 82
5.6.5 Influence of inference function 87
5.6.6 Influence of rejection level 88
5.6.7 The necessary number of simulations 89
5.6.8 Validation of simulated heads 92
5.7 Summary 94

6 Case B: Calibration and uncertainty estimation in a regional aquifer system 97
6.1 Conceptual model 97
6.1.1 Geology 98
6.1.2 Recharge 99
6.1.3 Streams 99
6.2 Observation data 100
6.3 Monte Carlo simulations 100
6.4 GLUE analysis 103
6.4.1 Head results 103
6.4.2 Stream discharge results 106
6.4.3 Parameter estimates 107
6.4.4 Reducing the GLUE rejection criteria 108
6.4.5 Convergence 109
6.5 Summary and conclusion 111

7 Summary and conclusion 113
7.1 Uncertainty estimation 113
7.2 The classical regression approach 115
7.3 The GLUE methodology 115

7.3.1 Synthetic case study 116
7.3.2 Regional aquifer case study 118
7.4 Remarks on the GLUE methodology 119

A Reliability methods **121**
A.1 Reliability 121
A.2 The Monte Carlo Method 124
A.3 The Latin Hypercube Method 126
A.4 First Order Reliability Method 127
A.4.1 Linear Safety Margin 127
A.4.2 Geometrical Interpretation of the Reliability Index 129
A.4.3 Non-linear Safety Margin 130
A.4.4 Sensitivity 132

B Simulation of random variables **135**
B.1 Generation of random numbers 135
B.2 Inverse Method 136
B.3 Box-Muller method for simulating normal distributed numbers . 137
B.3.1 Simulation of Log-normal distributed numbers 137
B.4 Simulation of correlated variables 138
B.4.1 Simulation of stochastic fields 139

List of Symbols

In general, lower case **bold** type represents vectors, e.g. $\mathbf{a}$, upper case **bold** type (e.g. $\mathbf{A}$) represents matrices, sans serif typed letters and upper case Greek typed letters, e.g. x and Θ, represent stochastic variables. In other words, letters representing a vector of stochastic variables are **bold** and sans serif formatted: e.g. $\mathbf{x}$. The page on which a symbol is first introduced is given in brackets.

Roman symbols

$a_{i,j}$	:	distance between cell nodes	[20]
$a_0, a_1, ..., a_N$	:	constants	[127]
A	:	area	[18]
$\mathbf{A}$	:	coefficient matrix	[24]
A_j	:	wetted cross section area in cell j	[21]
$\mathbf{A^k}\left(\hat{\boldsymbol{\theta}}^{\mathbf{k}}\right)$	:	$\left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]^T\left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]$	[20]
$A_{i,j}$	:	area of the wetted interface between cell i and cell j	[21]
$A_{poly,i}$	:	area of the i^{th} Thiessen polygon	[20]
$A_{river,j}$	:	area of the water surface within cell j	[22]
$\mathbf{b}$	:	constant vector	[24]
$b_{i,j}$	:	length of polygon side	[20]
$b_{x,j}$	:	cell width in x-direction at nodal point of cell j	[28]
$b_{y,j}$	:	cell width in y-direction at nodal point of cell j	[28]
c	:	concentration	[25]
$C_{i,j}$	:	conductance between cell i and cell j	[20]
C_{θ_i,θ_j}	:	covariance between θ_i and θ_j	[50]
$\mathbf{C}$	:	covariance matrix	[41]
$\mathbf{C}_{\boldsymbol{\psi}^*}$	:	expected error covariance of the observed state variables	[44]
$\mathbf{C}_{\boldsymbol{\theta}^*}$	:	covariance of the prior parameter estimate	[45]
$\mathbf{C}_{z^*}$	:	$\begin{pmatrix}\mathbf{C}_{\boldsymbol{\psi}^*} & 0\\ 0 & \mathbf{C}_{\boldsymbol{\theta}^*}\end{pmatrix}$	[45]
$\mathtt{CAP}_i$	:	the i^{th} capture zone	[76]
d	:	distance	[73]
$\mathbf{d}^k$	:	Gauss-Newton direction vector	[73]
$D_{i,j}$	:	hydrodynamic dispersion	[25]

$dz_{i,j}$	:	effective height of interface between cell i and cell j	[20]
dz_i	:	height of cell i	[20]
$f_{\psi^*\|\theta}(\psi^*\|\theta)$	:	joint conditional probability density function of the observations	[46]
g	:	acceleration due to gravity	[18]
$g(\cdot)$	:	failure function	[122]
G	:	solute sink/source term	[25]
I	:	integral scale	[73]
$I(\cdot)$	:	indicator function	[73]
i, j, k	:	indexes	[19]
J	:	hydraulic gradient	[36]
$J(\boldsymbol{\theta}^*)$	:	least square error	[44]
$J_D\left(\hat{\boldsymbol{\theta}}^k\right)$	:	Jacobian matrix of state variables, $\boldsymbol{\psi}$, with respect to parameters, $\boldsymbol{\theta}$	[48]
J_v	:	vertical hydraulic gradient	[36]
$K1, \cdots, K7$	:	horizontal conductivity in zones 1 to 7.	[102]
K_h, K_v	:	horizontal and vertical conductivity	[20]
K_x, K_y, K_z	:	hydraulic conductivity in x, y and z-direction	[17]
$L(\psi_i^*\|\boldsymbol{\theta})$	:	likelihood of ψ_i^* given $\boldsymbol{\theta}$ (point likelihood)	[57]
$L(\boldsymbol{\psi}^*\|\boldsymbol{\theta})$	:	likelihood of $\boldsymbol{\psi}^*$ given $\boldsymbol{\theta}$ (global likelihood)	[57]
$L(\boldsymbol{\theta})$	:	prior likelihood function	[64]
$L_p(\boldsymbol{\psi}^*\|\boldsymbol{\theta})$	:	posterior likelihood of $\boldsymbol{\psi}^*$ given $\boldsymbol{\theta}$	[64]
L_{Ψ^*}	:	likelihood of Ψ^*	[58]
M	:	Manning number	[18]
M	:	safety margin	[127]
N_{acc}	:	number of accepted simulations	[90]
$N_{cell,j}$	:	number of cells interacting with cell j	[19]
N_{int}	:	number of intervals	[126]
N_{obs}	:	number of observations	[48]
N_{par}	:	number of parameters	[48]
N_{rep}	:	number of repetitions	[127]
N_{sim}	:	number of simulations	[76]
P	:	wetted parameter	[18]
$\hat{P}(x_{0,i}, y_{0,i})$	:	probability estimate that a given particle starting in $x_{0,i}, y_{0,i}$ will be captured in the well	[90]
$P(x_{0,i}, y_{0,i})$	:	true probability that a given particle starting in $x_{0,i}, y_{0,i}$ will be captured in the well	[90]
P_f	:	failure probability	[122]
q	:	Darcy/filter velocity	[32]
q_b	:	specific discharge	[40]
Q	:	discharge	[17]
$Q_{i,j}$	:	flow between cell i and cell j	[19]
Q_{out}	:	sink/source term	[17]

$Q_{out,j}$	:	sink/source term from cell j	[19]
$Q^*_{i,j}$	:	correction flow term	[23]
r	:	resistance	[121]
R	:	stochastic resistance variable	[122]
R	:	hydraulic radius	[18]
R_j	:	hydraulic radius in cell j	[21]
$RCH1, RCH2$	:	recharge	[102]
s	:	load	[121]
s_h	:	calculated standard deviation on observed heads due to ignored heterogeneity	[37]
s_{q_b}	:	calculated standard deviation on q	[40]
s_{ψ_l}	:	calculated standard deviation on the l^{th} state variable ψ_l	[50]
s_r	:	calculated standard error of the regression	[50]
s_{θ_i}	:	estimated standard deviation on parameter estimate θ_i	[49]
$s(x_{0,i}, y_{0,i})$	:	standard error of the probability estimate, $\hat{P}(x_{0,i}, y_{0,i})$	[90]
S	:	stochastic load variable	[123]
S_f	:	friction slope	[18]
S_0	:	bottom slope	[18]
$S_{1,j}$	:	storage capacity in cell j at the beginning of timestep	[22]
$S_{2,j}$	:	storage capacity at the current iteration	[22]
S_y	:	specific yield	[23]
S_s	:	specific storage coefficient	[17]
t	:	time	[17]
$t(\cdot)$	:	Student-t distribution	[49]
TU	:	transmissivity	[102]
u	:	standard normal distributed variable	[128]
$\mathbf{u}$	:	vector of standard normal distributed variables	[129]
v_i	:	flow velocity in i direction (Darcy velocity in sub-surface flow)	[25]
$v_{i,j}$	:	flow velocity between cell i and cell j (Darcy velocity in sub-surface flow)	[25]
v_x, v_y, v_z	:	flow velocity in x, y and z direction	[26]
$v_{x,j}, v_{y,j}, v_{z,j}$	:	flow velocity in x, y and z direction in cell j	[27]
$\mathbf{V}$	:	weight matrix	[41]
w_f	:	failure space	[122]
w_s	:	safe space	[122]
x, y, z	:	spatial coordinates	[17]
y	:	water depth	[17]
$\mathbf{z}$	:	$= (\boldsymbol{\psi}, \hat{\boldsymbol{\theta}})^T$	[45]
$\mathbf{z}^*$	:	$= (\boldsymbol{\psi}^*, \boldsymbol{\theta}^*)^T$	[45]
z_b	:	bottom level	[18]
z_i	:	node level of cell i	[20]
z_i^-	:	bottom level of cell i	[20]
z_i^+	:	top level of cell i	[20]

Greek symbols

α	:	significance level	[74]
$\alpha_{i,j}$	:	dispersivity in the i-direction due to flow in the j-direction	[26]
α_l	:	correlation length	[37]
$\boldsymbol{\alpha}$	:	vector perpendicular to the failure surface in the β-point.	[129]
β	:	reliability index	[128]
ΔH	:	difference between the minimum and maximum value in a time series of potential head	[38]
Δsto_j	:	storage term in cell j	[19]
θ_i	:	the i^{th} parameter	[44]
$\boldsymbol{\theta}$	:	vector of parameters	[44]
$\boldsymbol{\theta}^*$	:	vector of expected parameters	[45]
$\hat{\boldsymbol{\theta}}$	:	vector of estimated parameters	[45]
θ_{eff}	:	effective porosity	[25]
Θ	:	stochastic parameter	[125]
$\boldsymbol{\Theta}$	:	vector of stochastic parameters	[123]
μ	:	mean value	[73]
μ_{M}	:	mean value of safety margin	[127]
ρ^k	:	scalar step size	[51]
$\rho(d)$	:	spatial correlation	[73]
σ	:	standard deviation	[73]
σ_0^2	:	weighted variance of the observations	[57]
σ_ϵ^2	:	weighted variance of the residuals	[57]
σ_{lnK}	:	standard deviation of the log-transformed (natural) hydrological conductivity	[37]
σ_{M}	:	standard deviation of the safety margin	[127]
σ_r^2	:	estimated common error variance of the regression	[50]
σ_{topo}	:	standard deviation of the topography within a computational grid	[38]
$\omega_{(\cdot)}$	:	weight	[50]
$\Phi(\cdot)$	:	standard normal distribution function	[128]
ψ	:	water level or groundwater hydraulic head	[18]
Ψ	:	a subset of the set of possible observed values for which membership is gradual	[58]
ψ^n	:	vector of water level or groundwater hydraulic head at time step n	[24]
ψ_i	:	water level or groundwater hydraulic head in cell i	[20]
ψ_j	:	water level or groundwater hydraulic head in cell j	[20]
$\psi_i(\boldsymbol{\theta})$	:	computed water level or groundwater hydraulic head given parameter vector $\boldsymbol{\theta}$	[44]
$\boldsymbol{\psi}^*$	:	vector of observed state variables	[44]

Acronyms

FORM	:	First Order Reliability Method	[5]
GLUE	:	Generalized Likelihood Uncertainty Estimations methodology	[6]
JPDF	:	joint probability density function for a number of random variables	[4]
PDF	:	likelihood density function	[8]
MC	:	Monte Carlo	[5]
PDF	:	probability density function	[4]
RCZ	:	reference capture zone	[80]

Summary

The data basis on which groundwater models are constructed is in general very incomplete, and this leads to uncertainty in model outcome. Groundwater models form the basis for many, often costly decisions and if these are to be made on solid grounds, the uncertainty attached to model results must be quantified. This study was motivated by the need to estimate the uncertainty involved in groundwater models.

Chapter 2 presents an integrated surface/subsurface unstructured finite difference model that was developed and applied to a synthetic case study.

The following two chapters concern calibration and uncertainty estimation. Essential issues relating to calibration are discussed. The classical regression methods are described; however, the main focus is on the Generalized Likelihood Uncertainty Estimation (GLUE) methodology. The next two chapters describe case studies in which the GLUE methodology was applied.

Capture zone modelling was conducted on a synthetic stationary 3-dimensional flow problem involving river, surface and groundwater flow. Simulated capture zones were illustrated as likelihood maps and compared with a deterministic capture zones derived from a reference model. The results showed that the reference capture zone was predicted within the 95% prediction zone. The results depended on a subjective criterion for model fit. However, it was argued that it would be advantageous to base the criterion for model fit on studies of expected errors in prediction.

The GLUE methodology was applied to a regional aquifer system that had previously been the target of parameter and uncertainty estimation within the classical regression framework. The prediction intervals achieved with regard to heads and stream flows were validated against observations. In comparison with the results from the regression analysis, a high number of observations were found to fall outside the 95% prediction bounds. The spatial distribution of the observation points where validation failed indicated errors in the conceptual description of the aquifer system.

Summary in Danish

Datagrundlaget, hvorpå grundvandsmodeller bygger er ofte meget ufuldstændigt, og dette bevirker, at resultaterne, der fremkommer fra disse modeller, er usikre. Grundvandsmodeller danner grundlag for mange og ofte bekostelige beslutninger. Hvis disse beslutninger skal tages på et solidt grundlag er det nødvendigt, at kvantificere usikkerheden knyttet til grundvandsmodellerne. Nærværende studium er motiveret i behovet for usikkerhedsestimering inden for grundvandsmodellering.

Kapitel 2 præsenterer en integreret overflade/grundvandsmodel byggende på et ustruktureret finit differens beregningsnet. Modellen er udviklet under projektforløbet og anvendt på et syntetisk setup.

De efterfølgende to kapitler omhandler kalibrering og usikkerhedsestimering. Centrale emner i relation til kalibrering er diskuteret. De klassiske regressionsmetoder er beskrevet, men fokus er sat på Generalized Likelihood Uncertainty Estimation (GLUE) metodologien. De næste to kapitler beskriver to case-studier, hvor GLUE-metodologien er anvendt.

Bestemmelse af infiltrationsopland er udført på et syntetisk 3-dimmensionalt strømningsproblem omfattende vandløb, overflade og grundvandsstrømning. De beregnede infiltrationsoplande er illustreret som likelihood-kort. Resultaterne viste at reference-infiltrationsoplandet kunne bestemmes inden for 95% prediktionszonen. Resultaterne var afhængige af de subjektive kriterier, der anvendes til beskrivelse af model-fit. Der bliver dog argumenteret for, at man med fordel kan beskrive model-fit kriterierne ud fra studier af den forventede fejl i prediktionerne.

GLUE-metodologien blev anvendt på et regionalt aquifersystem, der tidligere, vha. regressionsanalyse, havde have været genstand for parameter- og usikkerhedsestimering. Prediktionsintervaller for trykniveauer og vandløbsafstrømning blev valideret mod observationer. Et, i forhold til regressionsanalysen, højt antal trykniveauer befandt sig uden for 95% prediktionsgrænserne. Den rumlige fordeling af observationspunkterne, hvor valideringen slog fejl indikerede fejl i den konceptuelle beskrivelse af aquifersystemet.

CHAPTER 1

Introduction

Groundwater is the predominant source of water supply in Denmark. Groundwater is used for human consumption, industrial processes and agriculture. More then 99% of all water for human consumption in Denmark originates form groundwater. Only in a few areas do the problems relating to groundwater in Denmark concern the *quantity* of groundwater. In total only 25 % of the estimated Danish resource is utilised (Danmarks Statistik 2002). On the other hand, the *quality* of groundwater in Denmark has been the focus of considerable attention. An increasing number of wells have been closed due to manmade pollution. There are many sources of pollution, but nitrate, pesticides and MTBE pollution (gasoline additive) are the main reason for closing abstraction wells. In Denmark the amount of fertiliser and pesticides has been decreasing since the mid 1980s and many pesticides have been forbidden. But since most of the drinking water is 20 years old or more, there is a pronounced risk that we have only seen the beginning of nitrate and pesticide problems, and that these problems will increase for some time to come.

The water treatment at the majority of waterworks in Denmark is simple, consisting only of mechanical filtering and oxydation.

This means that the quality of drinking water depends on the reduction of solutes, e.g. nitrate and pesticides, in groundwater (Miljøstyrelsen 2000).

Groundwater models are used as a tool to regulate land use and to prioritise the remediation of polluted sites. The cost of ensuring unpolluted groundwater is high, both for the authorities and for the industries and farmers affected by the regulations. Decisions must therefore be based on reliable models!

1.1 Application of groundwater models

Groundwater models play an important role in the assessment of both the quality and quantity of groundwater. Models have been used for problems ranging from overall water balance calculations to small-scale pollution problems, including chemical and biological reactions. The status of the various groundwater modelling systems currently applied is presented in Table 1.1. After (Refsgaard 1996)

Table 1.1: Status of current of hydrological modelling systems. After Refsgaard (1996)

Status of application	Field of problem		
	Groundwater resources assessment	Groundwater pollution	
		Point sources	Non-point sources
Adequacy of Scientific Basis	Good	Good	Fair
Scientifically Well Tested ?	Good	Good	Fair
Validation on Pilot Schemes ?	Adequate	Partially	Very limited
Practical Applications	Standard/Part	Standard/Part	Very limited
Major Constraint for Practical Applications	Administrative	Techno/Admin	

LEGEND:

Adequacy of Scientific Basis

- Poor: Large and crucial needs for improvements in scientific basis
- Fair: Considerable needs for improvements in scientific basis
- Good: Some needs for improvements in scientific basis
- Very good: No present significant need for improvements in scientific basis

Scientifically Well Tested ?

- Poor: Large needs for fundamental tests of scientific method
- Fair: Considerable needs for testing (some) of the scientific basis
- Good: Some needs for testing of the scientific basis
- Very good: No present significant need for testing of the scientific basis

Validation on Pilot Schemes ?

- Nil: No successful validation on well controlled pilot schemes so far - urgent need for validation on pilot schemes
- Very limited: A few (a couple of) validation cases - considerable needs for more validation projects on pilot schemes
- Partially: Some cases with successful validation on pilot schemes - some needs for further validations
- Adequate: Many good validations - no further present needs

Practical Applications

- Nil: Practically no operational applications
- Very limited: a few well proven cases of operational practical applications
- Some cases: Some cases of well proven a few well proven cases of operational practical applications
- Standard/Part: Standard professional tool in some regions
- Standard: Standard professional tool in many regions of the world

Major Constraint for practical applications operational practical applications

- Data: Data availability a major constraint
- Science: Inadequate scientific basis is a major constrain
- Technology: A technology push is required in order to make well proven methods more widely applicable
- Administrative: Administrative tradition or missing economic motivation is a major constraint

1.1.1 Water quantity modelling

Groundwater models are used to evaluate the effect of exploiting groundwater resources. A change in the location of abstraction sites or in the amount of water abstracted will influence the local and perhaps the regional water balance, leading for example to changes in groundwater level, changes in river discharges and/or changes in evapotranspiration. The effects on water balance are often indirect. For example a lowering of the groundwater level might result in subsidence damage to houses, or problems in water quality from chemical reactions that have been developed in the drained region of the saturated zone. A reduction in river discharge might alter the habitat of certain organisms and affect the feeding, and therefore the quality, of receiving waters. In Denmark groundwater model applications have traditionally dealt with these water balance issues, and as the table shows, the reliability of these kinds of applications is therefore high.

1.1.2 Water quality modelling

Over the last one or two decades the focus has shifted towards studying how groundwater can be protected, with the main emphasis on modelling capture zones and transport time. Capture zone modelling has become a valuable tool in planning future land use and in prioritising the remediation of polluted sites. However, modelling of the groundwater transport of nitrate, pesticides and other solutes has yet to become standard practice, Table 1.1.

Given the tendency for model applications to model more and more processes on a smaller and smaller scale, both the amount of data required for model construction, and the variety of data types used, are constantly increasing. There is also an ever greater need to discretise the data employed. When working with a large-scale problem the effect of small-scale parameter variations may be neutralised, but when the focus turns to problems on a smaller scale we can no longer rely on the averaging process. Local variations in soil properties and groundwater recharge have often been neutralised in catchment water balances, but such variations may have a significant influence on e.g. capture zone estimates. No matter what the scale of the groundwater models implemented, it is necessary to calibrate the key parameters, and the amount of data that can be used for calibration increases only slightly as more and more processes are included. The result is greater uncertainty - often seen in terms of poor validation.

As the previous two subsections indicate, the reliability of modelling results depends on the application. In general reliability can be expected to decrease as the process scale under consideration decreases - thus groundwater resource simulations are likely to be more certain then small-scale pollution simulations where dispersion and chemical and biological processes are important.

Definitions

A (groundwater) **modelling system** is defined as a set of physical processes that convert input variables and parameters into output variables.

A **conceptual model** is a simplified description of physical conditions in the model area and of the processes that are to be modelled. The term **conceptual model error** refers to errors in the model outcome that originate from errors in the conceptual description of the problem under consideration.

A **groundwater model**, or **model** in short, is the numerical formulation of the conceptual model. The term **model error** refers to errors in the model outcome that originate from both errors in the conceptual description and errors in the numerical solution of the problem under consideration.

A **geological model** is a component of the conceptual model describing geological formations, often in a simplified manner, as layers and lenses.

A **variable** is characterised as being measurable and time-varying, whereas a **parameter** is characterised as being time-independent.

System input parameters and variables represent model input, and **system state variables** represent model outcome, such as potential head, water level, concentration, etc.

Prior statistics are the probability density functions (pdf) with related parameters that are present prior to the modelling. E.g. log-normal pdf with mean and standard deviation of hydraulic conductivity, Gaussian pdf of average net-precipitation, a number of geological models with equal probability.

Posterior statistics are prior statistics updated according to new knowledge archived during modelling.

1.2 Calibration and uncertainty estimation

The uncertainty of the outcome of a groundwater model can be analysed by propagating the different sources of uncertainty through the model to the model output. As Melching (1995) has pointed out, the ideal situation is to have the joint probability distribution function (JPDF) for the significant sources of uncertainty. It is, however, very difficult or impossible to determine the statistics of all models, basic parameters and variables, and crude assumptions usually have to be made. In fact the challenge in uncertainty estimation often lies in estimating the input statistics (probability density function (PDF) and correlation). It is generally accepted that groundwater models have to be calibrated in order to reproduce the system that is going to be modelled. This is a direct result of the difficulties in establishing an error-free conceptual model and good prior information on all key variables and parameters.

Reliability estimation methods can be roughly categorised into three types: 1) forward stochastic methods, 2) conditioning stochastic methods and 3) regression

Note

Parameter estimation is an often used synonym for calibration. Calibration is however not restricted to estimation of parameters: all unknown **models**, **parameters** and **variables** can be the target of estimation. To emphasise this, unknowns are in general, throughout this thesis, referred to as **models**, **parameters** and **variables**. Occasionally, **parameters** is used in short for all types of unknowns.

methods, see Fig. 1.1.

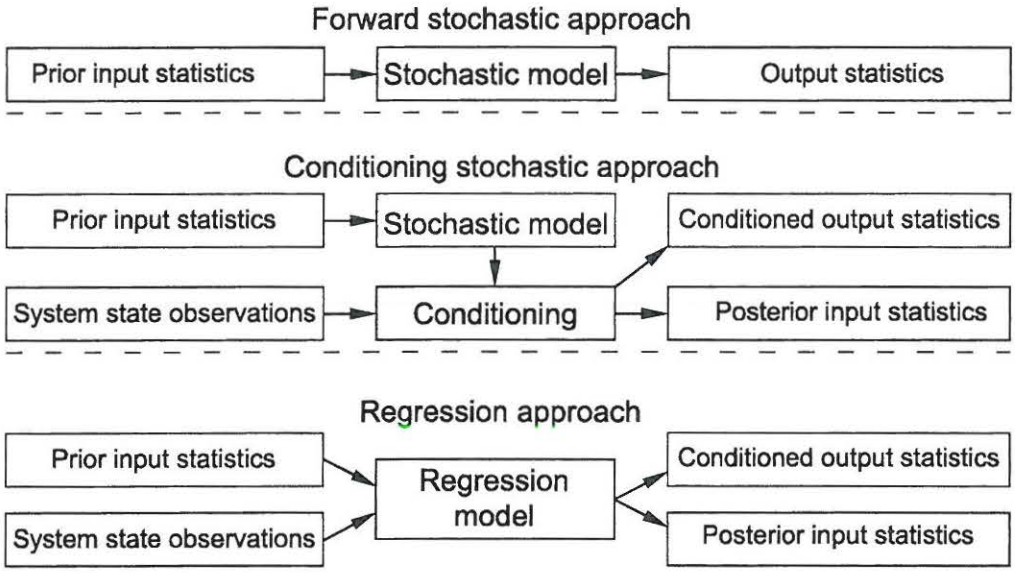


Figure 1.1: Categories of approach to reliability estimation.

1) Forward stochastic approach: Output statistics are estimated on the basis of prior input models, variables and parameters. The term *forward* refers to the fact that there is no updating of the prior information included in this process. Freeze (1975) considered the porous media in a synthetic one-dimensional flow regime as a stochastic set of macroscopic elements in which hydrological conductivity, compressibility and porosity were defined by probability density functions. The Monte Carlo method (Section A.2) was used to estimate the statistics of the hydraulic heads.

Another forward stochastic method is the First Order Reliability Method, FORM (Section A.4) which used a first-order Taylor series expansion at a specific linearisation point in order to calculate a measure of reliability, e.g. the probability of a concentration exceeding a limit value. Sitar et al. (1987) and Wu and Cawlfield (1992) have demonstrated how FORM can be applied to groundwater flow and transport problems.

Definitions

Confidence intervals are a measure of the uncertainty that originates from having a limited number of observations in the prediction of parameter values and state variables. The confidence intervals will diminish as the number of observations increases.

Prediction intervals are a measure of the overall uncertainty, including contributions from the estimation of parameters, from numerical errors and from conceptual model errors.

An **aquifer** is "*a geological formation, or a group of formations, which contains water and permits significant amounts of water to move through it*" (Bear 1979). **Groundwater reservoir** is another frequently used term.

An **aquitard** is "*a geological formation which is of a semipervious nature; it transmits water at a very low rate compared with the aquifer*" (Bear 1979).

2) Conditioning stochastic approach: The valuable measurements of system state variables are used for conditioning of prior statistics on models, parameters and variables, and thereby also the model outcome. The prior input statistics are updated and posterior input statistics are obtained.

The Monte Carlo-based Generalized Likelihood Uncertainty Estimations methodology, GLUE, proposed by Beven and Binley (1992), is one example of a conditioning stochastic method. In the GLUE methodology a wide variety of combinations of models, parameters and variables are taken as possible simulators of the system until the opposite is proven. In short, a large number of possible combinations of models, parameters and variables are evaluated and weighted according to a given model-fit criterion. See Chapter 4.

Feyen et al. (2002) used a Bayesian approach to stochastic capture zone delineation, incorporating both transmissivity measurements and head observations.

Neuman (2002) suggested a strategy accounting for conceptual model uncertainty using the maximum likelihood Bayesian model averaging approach. Conceptual models and their parameters are conditioned on the basis of available system state observations.

3) Regression approach (inverse approach) This type of modelling approach is at present the one most commonly used to estimate output reliability. On the basis of observed system state variables an optimal parameter set is sought by minimising a certain model fit criterion. The regression analysis provides approximate confidence, together with prediction intervals of system state variables and confidence intervals for the estimated parameters (Cooley 1977; Yeh 1986; Hill 1992; Christensen and Cooley 1999 and Section 3.5). Carrera and Neuman (1986a) propose the Maximum Likelihood framework including prior information for the

inverse problem, see Section 3.6.

Example 1.1 is an attempt to describe the three different types of approach described above in relation to a simple groundwater problem.

Example 1.1 *Stationary groundwater flow in an artesian homogeneous aquifer is considered. The flow in the aquifer is assumed to be one-dimensional and a semi-permeable aquitard is present on top of the aquifer from which groundwater is leaching. A water divide constitutes the left boundary (no-flow boundary) and a pressure boundary is formulated at the symmetry line of the river. Head potential in the aquifer is recorded in wells no. 2 and 3, ψ_2^* and ψ_3^*, and the head potential in well no. 1, ψ_1, is to be estimated. The aquifer inflow to the river, Q_r, is estimated from river discharge observations. Fig. 1.2.*

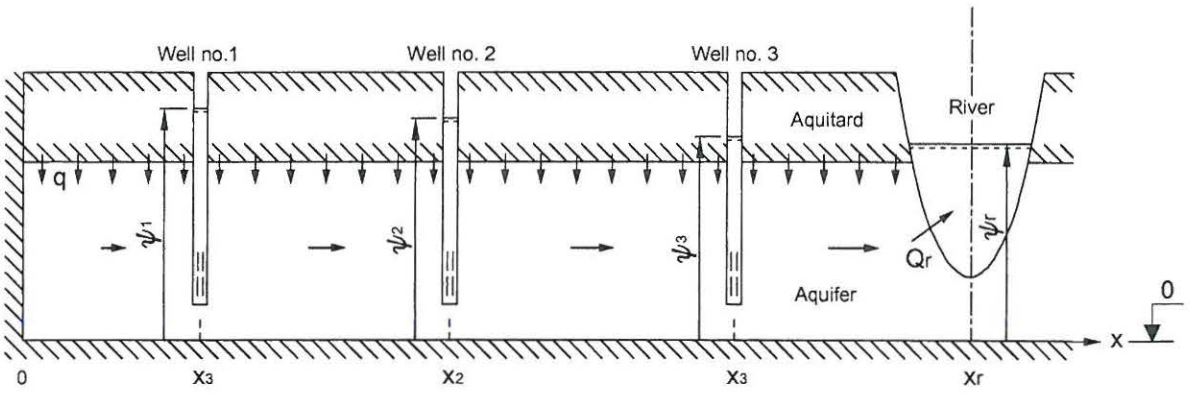


Figure 1.2: Schematic cross section view of the groundwater flow problem

The head potential in the aquifer can be described by the differential equation

$$\frac{\partial^2 \psi}{\partial x^2} T + q = 0 \tag{1.1}$$

And if the boundary conditions are simplified to

$$\begin{aligned} \psi(x_r) &= \psi_r \\ \tfrac{d\psi(0)}{dx} &= 0 \end{aligned} \tag{1.2}$$

the solution to Eq. 1.1 becomes

$$\psi(x) = \frac{q}{2T}\left(x_r^2 - x^2\right) + \psi_r \tag{1.3}$$

Let us for simplicity assume that the boundary conditions and model are error free and that the only uncertainty in the estimate of the head potential originates from uncertainty in the infiltration, q, and the aquifer transmissivity, T. Prior to the simulation, q is assumed to be Gaussian distributed with μ_q and σ_q. T is assumed to be log-normal distributed with μ_T and σ_T.

In the ***forward stochastic approach*** *all state variable (head and river inflow) observations are ignored, and only the prior statistics on q and T are used in the estimation of ψ_1. If, for instance, the Monte Carlo method is applied to the problem a large number, N, of random realisations of q and T are made, $(q_1, ..., q_i, ..., q_N)$ and $(T_1, ..., T_i, ..., T_N)$, each resulting, through Eq. 1.1, in N head potential values, $(\psi_{1,1}, ..., \psi_{1,i}, ..., \psi_{1,N})$, in well no. 1. From a statistical analysis a discrete probability density function of ψ_1 can be found.*

If we follow the ***conditional stochastic approach*** *the uncertainty in ψ_1 is still estimated from the uncertainty of q and T, using e.g. Monte Carlo simulation as described above, and subsequently the estimate of ψ_1 and the prior estimates of q and T are conditioned, based on the observation of ψ_2^*, ψ_3^* and Q_r^*. The conditioning is introduced because the prior statistics on q and T are incomplete. Following the GLUE methodology, each realisation will be given a likelihood weighting, $L(q_i, T_i)$, according to how close the calculated head potential is to the observed value. The final outcome is a posterior density function for q, T and a likelihood density function (LDF) of ψ_1. See Chapter 4.*

In the ***regression approach*** *an objective function is formulated (e.g. the Least Square Objective Function or the Maximum Likelihood Objective Function) describing model fit. The standard objective function is a function of the observations ψ_2^*, ψ_3^* and Q_r^* and the corresponding computed values, $\psi_2(\hat{q}, \hat{T})$, $\psi_3(\hat{q}, \hat{T})$ and $Q_r(\hat{q}, \hat{T})$, where $\hat{q}$ and $\hat{T}$ are current estimates of q and T, see e.g. Eq. 3.4.*

Alternatively, prior information on q and T can be built into the objective function so that the solution is constrained by observation as well as by prior parameter information, see e.g. Eq. 3.5.

Based on the combination of $\hat{q}$ and $\hat{T}$, which minimises the objective function, the statistics on the estimate of $\hat{q}$ and $\hat{T}$ and an estimate of the PDF for ψ_1 can be obtained. See Section 3.8 and 3.9.

1.3 Objectives

The main objectives of this study are:

(i) To assess the uncertainty of *all* the state variables that are the target of the model application, rather than those used in model calibration alone.

(ii) To obtain approximate uncertainty estimates in relation to the whole problem, rather then obtaining exact uncertainty estimates in relation to only a part of the problem.

(iii) To evaluate uncertainty, starting from the premise that all input data are a priori uncertain, and that this uncertainty can be reduced by adding conditional data to the model.

Re: i: Most groundwater models are calibrated using stream flow and groundwater head observations and, as a result, are designed to predict stream flow and groundwater heads. However, it is only in certain situations that these quantities are interesting in themselves. Derived quantities, such as groundwater pore velocities and flow directions are more important, because these form the basis for transport and capture zone estimates. The fact that models are used for predicting state variables that have not been directly used in calibration will reduce the reliability of the model result.

The small predictive uncertainty of the state variables used in the calibration does not guarantee that the predictive uncertainty of the state variables that have *not* been used in the calibration will also be small. It is therefore insufficient to describe the overall uncertainty in terms of the uncertainty of the state variables used in the calibration alone. All uncertainty must however be transmitted to the state variable of interest.

Re: ii: The uncertainty of the predicted state variables originates from a number of sources, and the aim of modelling uncertainty must be to account for all the major sources. My argument is based on the general thesis formulated by C.A. Cornell and quoted by Melching (1995):

"It is important in engineering applications that we avoid the tendency to model only those probabilistic aspects that we think we know how to analyze. It is far better to have an approximate model of the whole problem than an exact model of only a portion of it"

In groundwater modelling this indicates that the traditional methods based on classical regression statistics can no longer be usefully applied and that more subjective measures of uncertainty have to be adopted.

Re: iii: In a deterministic approach the model is assumed to be certain until the opposite is proven, either by sensitivity testing or through obtaining additional knowledge, experience and data. The burden of proof should however be reversed, so that the *certainty* of the model must be proven. This principle is illustrated Fig. 1.3.

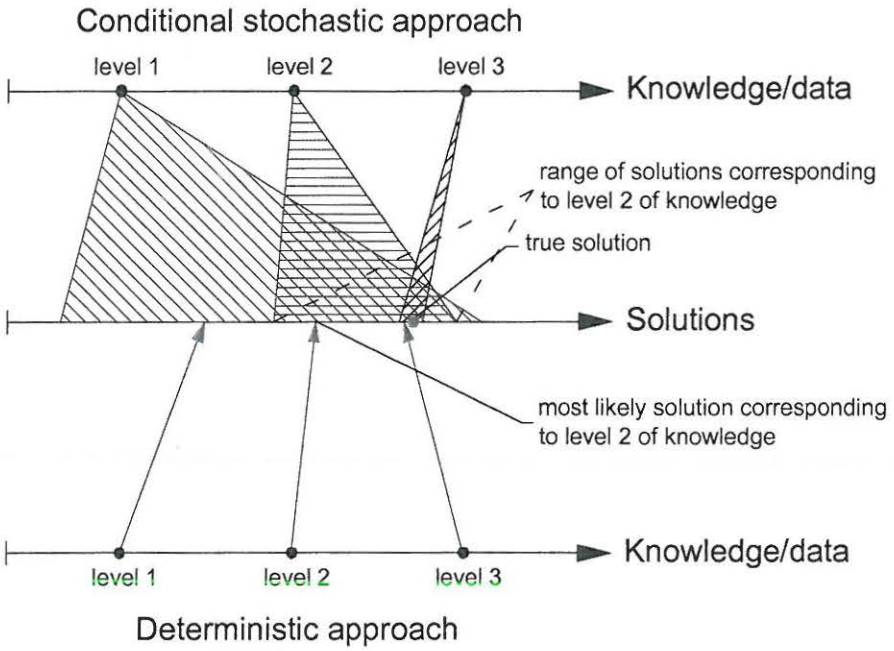


Figure 1.3: Deterministic versus conditional stochastic modelling approach.

Level 1 of knowledge results in a large range of possible solutions, including, it is hoped, the correct one. As the level of knowledge/data increases, the range of possible solutions narrows, until finally the deterministic approach and the conditional stochastic approach converge towards the same solution. This approach takes into account the general conditions during the simulation of groundwater - namely that the knowledge/data basis is far from being complete. Thus the oft-used phrase

"On the basis of current knowledge, experience and data, this is our best estimate of the true solution"

would be replaced with

"On the basis of current knowledge, experience and data, this is out best estimate of the possible solutions"

1.4 Outline of the thesis

Chapter 2 gives an introduction to the field of groundwater modelling. It presents both the governing equations and the integrated surface/subsurface unstructured finite difference model developed in the course of research for this ph.d.

Chapter 3 gives an introduction to traditional inverse calibration. General issues concerning calibration, including the effects of scale and nonlinearity, are discussed. The numerous sources of uncertainty are described. The Least Square and Maximum Likelihood Parameter Estimation methods are presented. The issue of multiple objective functions is briefly discussed. Finally, the chapter presents parameter and state variable statistics for Gaussian linear models.

Chapter 4 gives an introduction to the Generalized Likelihood Uncertainty Estimation methodology (GLUE).

Chapter 5 gives a presentation of uncertainty estimation in relation to capture zone modelling in a synthetic setup, where the GLUE methodology is adopted.

Chapter 6 shows how the GLUE methodology can be applied to a regional aquifer system. The results are compared with those obtained from an earlier study based on the Least Square Regression method.

Chapter 7 concludes the thesis and sums up the main results and conclusions.

Appendix A gives an introduction to reliability methods, including, Monte Carlo simulation, Latin Hypercube simulation and the First Order Reliability Method.

Appendix B concerns the important issue of generating random numbers and random fields.

CHAPTER 2

Groundwater modelling

The term "groundwater modelling" is used in this thesis as a general term for hydrological modelling with a special focus on groundwater problems. The term is often used to cover the modelling of several different hydrological processes such as groundwater flow, evapotranspiration, infiltration, surface flow, stream flow, drainage flow, etc. Various approaches have been applied to the different groundwater models in order to take into account these different hydrological processes. The complexity of the process integration ranges from using explicitly formulated sink/source terms and boundary conditions, as implemented in MODFLOW (McDonald and Harbaugh 1988), to developing fully implicit integrated models that simultaneously solve equations for unsaturated flow, saturated flow river flow, etc. MIKE SHE (Abbott et al. 1986) is one example of the latter type, which is known as an integrated hydrological model.

Various model types have been used for groundwater modelling - ranging from *empirical models* without any physical relation between input parameter/variables and output parameter/variables (e.g. artificial neural network models, Coulibaly et al. (2001)), through *lumped conceptual models*, which are semi-empirical with a physical basis (e.g. the rainfall-runoff model NAM (Nielsen and Hansen 1973)) to *distributed physically-based models* where some or all of the hydrological processes involved are described directly through the governing equations relating to these processes.

In principle distributed physically-based models have the capability (depending on data) to give a detailed description of hydrological processes. Groundwater head and flows can be resolved on a small scale depending on the numerical discretisation of the governing equations. MIKE SHE, IHDM (Beven et al. 1987) and MODFLOW are examples of distributed physically-based models.

The present thesis focuses on these distributed physically-based models. MODFLOWP (non-linear regression version of MODFLOW, (Hill 1994)), was used in *Case B - Calibration and uncertainty estimation in a regional aquifer system* reported in Chapter 6, and as part of this thesis FLEXFLOW, an integrated surface/sub-surface finite difference model based on a unstructured grid, was developed.

The development of FLEXFLOW was motivated by the lack of flexibility in existing finite difference models with regard to numerical discretisation, representation of geological units, river discretisation and surface, river and groundwater flow interaction.

The rigidity in existing finite difference models may cause additional uncertainty in the model outcome, and it is hoped that some of this uncertainty is eliminated in FLEXFLOW.

FLEXFLOW was used in *Case A: Capture zone modelling in a synthetic setup*, which is presented in Chapter 5 and in Jensen and Schaarup-Jensen (2002).

The present chapter deals with the unstructured finite difference model developed as part of this thesis.

2.1 Numerical discretisation

Finite difference groundwater models are traditionally based on a number of adjoining horizontal rectangular grids. MIKE SHE is limited to quadratic grids, Fig. 2.1A, whereas MODFLOW has rectangular grids with variable dimensions, with the restriction that $dx_{i,j} = const$ for all j and $dy_{i,j} = const$ for all i, Fig. 2.1B. Based on the MODFLOW2000 model (Harbaugh et al. 2000), Mehl and Hill (2002) introduce a local grid refinement using shared nodes. Finite element models (i.e. FEFLOW (Koskinen et al. 1996)) typically have greater flexibility in the grid construction because triangular, Fig. 2.1C, or rhomboid grids, Fig. 2.1D, are easily implemented.

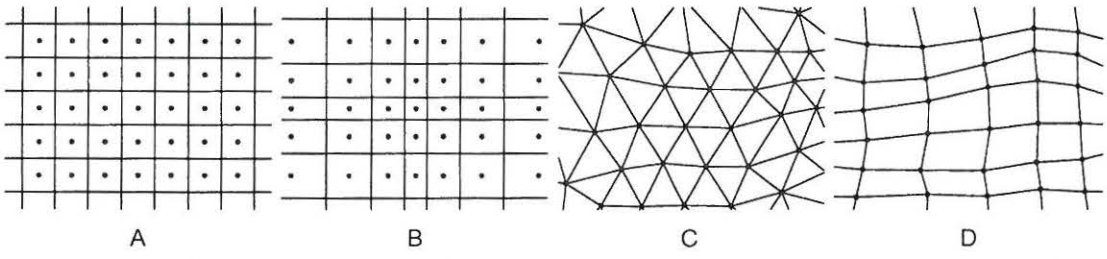


Figure 2.1: A) quadratic grids, B) rectangular grids with variable dimensions, C) triangular shaped grids, D) rhomboid grids

The integrated model presented here, FLEXFLOW, is a finite difference model with a flexible unstructured grid based on Thiessen polygons, Fig. 2.2. The

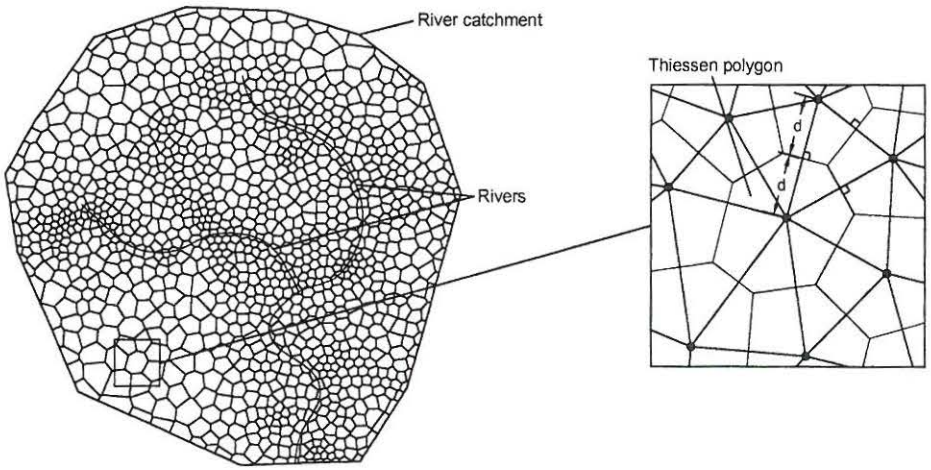


Figure 2.2: Example of horizontal discretisation of a river catchment

model has the flexibility of traditional finite element models and the simplicity of finite difference models when it comes to formulating the discretised equations to be solved. The grid can be refined for specified regions, e.g. near wells, rivers, etc., and adapts to rivers so that they retain their original placing.

2.2 Geological representation

In most groundwater models the user has the possibility of letting the vertical discretisation follow the geological layers. The number of layers is usually constant throughout the model.

FLEXFLOW works with the concept of geological units, which are a generalisation of the concept of geological layers and lenses. A unit is a spatial volume with a given deposit and contains layers as well as lenses. The vertical discretisation follows the geological units and a given unit may have a computational layer thickness of e.g. 5 m, while the neighbouring unit has a layer thickness of e.g. 30 m. In other words, calculation cells are placed where they are needed, Fig. 2.3.

2.3 Implementation of rivers

There are several traditional approaches to modelling rivers, ranging from models with simple boundary conditions on the top layer (MODFLOW), through mod-

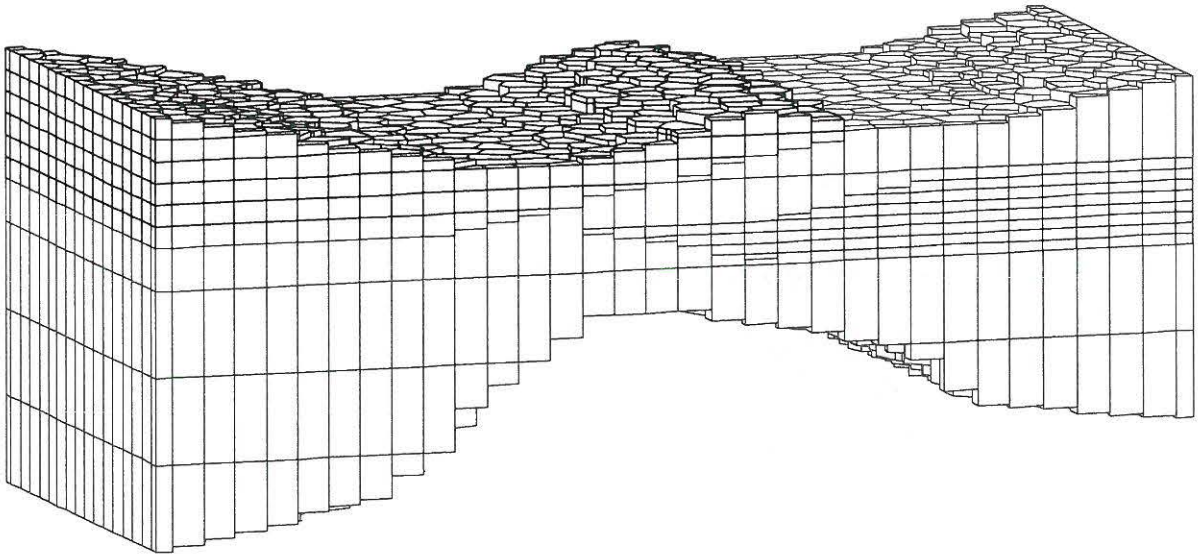


Figure 2.3: The FLEXFLOW *discritisation of three geological units, each with different vertical discritisation*

els in which separate groundwater and river models are dynamically coupled (MIKE SHE), to fully integrated surface/subsurface models (Waterloo model (Vanderkwaak and Sudicky 2000)). The river geometry is typically formed by a numerical grid. Thus in MIKE SHE rivers are placed in a zigzag pattern on the borderlines between cells, while in MODFLOW special river grids are assigned to the model. A detailed description of the alignment of rivers typically requires a very dense traditional grid or a flexible grid structure.

FLEXFLOW forms the numerical grid in accordance with the digitised river contours, and the river maintains its exact position, Fig. 2.2 and 2.4. In the case of wide rivers it is possible to have a two-dimensional river flow and water may freely flow from river cells to surface cells and vice versa, Fig. 2.4.

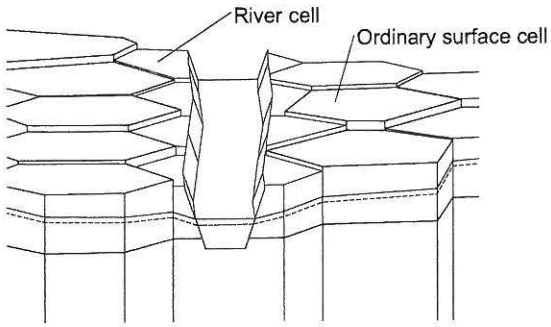


Figure 2.4: A 3-dimensional schematic representation of a river location in the numerical grid of FLEXFLOW

Two-dimensional surface flow, two-dimensional river flow and three-dimensional groundwater flow are implicitly fully integrated.

2.4 Governing equations

At the present stage two-dimensional surface flow and three-dimensional saturated sub-surface flow are modelled in FLEXFLOW.

2.4.1 Groundwater flow

The governing equation for 3-dimensional groundwater flow is given as in e.g. Bear (1972):

$$\frac{\partial}{\partial x}\left(K_x \frac{\partial \psi}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial \psi}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial \psi}{\partial z}\right) + -Q_{out} = \frac{\partial \psi}{\partial t} S_s \qquad (2.1)$$

where

ψ	groundwater potential [m]
K_x	hydraulic conductivities along x-axis [m s^{-1}]
K_y	hydraulic conductivities along y-axis [m s^{-1}]
K_z	hydraulic conductivities along z-axis [m s^{-1}]
Q_{out}	is a sink-source term [m^3 s^{-1} m^{-3}] representing all exchanges with the surroundings, such as water abstraction, drainage routing, groundwater-river exchange, etc.
S_s	is the specific storage coefficient [m^{-1}]

2.4.2 Surface flow

The governing equations for surface flow are the continuity and momentum equations. For one-dimensional flow the continuity equation is given as

$$\frac{\partial y}{\partial t} + \frac{\partial Q}{\partial x} = Q_{out}(x, y, t) \qquad (2.2)$$

where

Q	discharge [$m^2 s^{-1}$]
y	water depth [m]
O_{out}	sink/source term [m s^{-1}]

and the momentum equation is

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + gA\frac{\partial y}{\partial x} = gA(S_0 - S_f) \tag{2.3}$$

where

g	acceleration due to gravity [m s^{-2}]
A	wetted cross section area [m^2]
S_0	bottom slope [m m^{-1}]
S_f	friction slope [m m^{-1}]

According to diffusive wave theory the acceleration terms are neglected and the momentum equation becomes

$$\frac{\partial y}{\partial x} = S_0 - S_f \tag{2.4}$$

In the case of small bottom slopes the water level can be approximated to $\psi = z_b + y$, where z_b is bottom level and ψ is water surface level. The friction slope becomes

$$S_f \simeq \frac{\partial \psi}{\partial x} \tag{2.5}$$

The discharge based on the Manning formula and the wetted cross section area reads

$$Q = AMR^{\frac{2}{3}}S_f^{\frac{1}{2}} \tag{2.6}$$

where

M	Manning number
R	hydraulic radius, (A/P)
P	wetted perimetre

Eq. 2.6 and 2.5 applied to the continuity equation, Eq. 2.2, result in the governing equation for one-dimensional river flow under the diffusive wave approximation

$$\frac{\partial \psi}{\partial t} + \frac{\partial}{\partial x}\left(\text{sign}\left(\frac{\partial \psi}{\partial x}\right) AMR^{\frac{2}{3}}\left(\frac{\partial \psi}{\partial x}\right)^{\frac{1}{2}}\right) = Q_{out}(x,t) \tag{2.7}$$

where the sign $\left(\frac{\partial \psi}{\partial x}\right)$ function provides the sign of $\frac{\partial \psi}{\partial x}$.

In a similar manner the governing equation for two-dimensional surface flow can be derived as follows

$$\begin{aligned}\frac{\partial \psi}{\partial t} + \frac{\partial}{\partial x}\left(\text{sign}\left(\frac{\partial \psi_x}{\partial x}\right) A_x M_x R_x^{\frac{2}{3}}\left(\frac{\partial \psi_x}{\partial x}\right)^{\frac{1}{2}}\right) + \\ \frac{\partial}{\partial y}\left(\text{sign}\left(\frac{\partial \psi_y}{\partial y}\right) A_y M_y R_y^{\frac{2}{3}}\left(\frac{\partial \psi_y}{\partial y}\right)^{\frac{1}{2}}\right) = Q_{out}(x,y,t)\end{aligned} \tag{2.8}$$

2.5 Numerical formulation

The numerical formulation of the governing equations is based on Thiessen polygons cells. The model structure is analogous to that presented in the Thiessen multicell models (e.g. Bear (1979)) which were formulated in order to locate nodal points where observed heads were present. The present formulation is applied to standard cell scales in traditional finite difference or finite element models. To the best of the author's knowledge, the present formulation of the Thiessen polygon cells in the description of 3D dimensional flow is new. The numerical grid is auto-built from Delaunay triangles generated by the "Triangle" programme by Ruppert (1995). Fig. 2.5 shows the Thiessen polygonal cells.

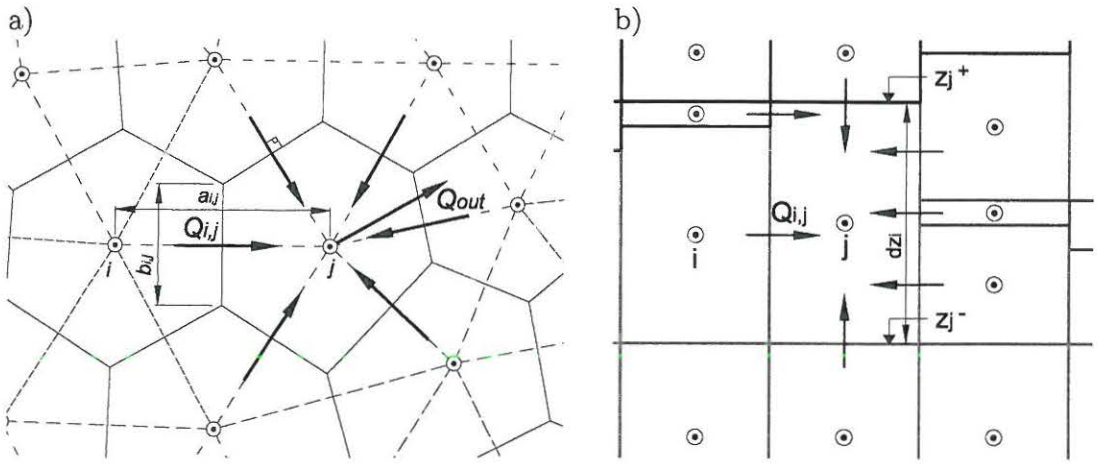


Figure 2.5: Thiessen polygonal cells - a) plan view, b) vertical cross section

The finite difference equations are derived from considering the water balance in the j^{th} numerical cell, Fig. 2.5, Eq. 2.9.

$$\left(\sum_{i=1}^{N_{cell,j}} Q_{i,j} - Q_{out,j} \right) \Delta t = \Delta sto_j \tag{2.9}$$

The number of cells, $N_{cell,j}$, with which a given cell interacts may vary from cell to cell. In the horizontal plane the number of interacting cells is often in the range of 6 to 12 cells, while only 1 or 2 cells interact in the vertical plane. where $Q_{i,j}$ is the i^{th} inflow to cell j, $Q_{out,i}$ is the sink source term and Δsto_j is the amount of stored water during the time step under consideration.

2.5.1 Potential flow terms

The difference between surface and sub-surface flow lies within the calculation of $Q_{i,j}$, or more precisely within the calculation of the conductance, $C_{i,j}$, between cell i and cell j. For all interacting cells $Q_{i,j}$ is given as

$$Q_{i,j} = (\psi_i^n - \psi_j^n)C_{i,j} \tag{2.10}$$

where ψ_i and ψ_j are water levels or potential head for surface cells and sub-surface cells respectively. The index n refers to the time step, where n is the time step at which the state variables are to be calculated.

Note that index i and j refer to the two connecting cells for which flow is considered. We do not distinguish between horizontal and vertical flow, so Eq. 2.10 describes the flow between all cells, both those that are horizontally and those that are vertically connected.

Conductance between two sub-surface cells

Darcian flow is assumed in the calculation of sub-surface flow. In the case of saturated flow in the horizontal plane the conductance becomes

$$C_{i,j} = \frac{2b_{i,j}dz_{i,j}}{a_{i,j}} \frac{1}{\frac{1}{K_{h,i}} + \frac{1}{K_{h,j}}} \tag{2.11}$$

where $K_{h,i}$ and $K_{h,j}$ are the horizontal conductivity in cells i and j, respectively, $a_{i,j}$ is the distance between cell i and j, $dz_{i,j}$ and $b_{i,j}$ is the effective height and width of the interface between cells i and j. See Fig. 2.5. In general $dz_{i,j}$ is given as

$$\begin{aligned} dz_{i,j} = 0.5(&\max\left[0, \min\left(\psi_i, z_i^+, z_j^+\right) - \max\left(z_i^-, z_j^-\right)\right] + \\ &\max\left[0, \min\left(\psi_j, z_i^+, z_j^+\right) - \max\left(z_i^-, z_j^-\right)\right]) \end{aligned} \tag{2.12}$$

where z_i is the node level and $z_i^- = z_i - dz_i/2$ and $z_i^+ = z_i + dz_i/2$ are the bottom level and top level of cell i, Fig. 2.5b.

The vertical conductance between two sub-surface cells is given as

$$C_{i,j} = A_{poly,i} \frac{1}{\frac{dz_i/2}{K_{v,i}} + \frac{dz_j/2}{K_{v,j}}} \tag{2.13}$$

where $A_{poly,i}$ is the area of the Thiessen polygon of cell i, and $K_{v,i}$ and $K_{v,j}$ are the vertical conductivities.

Conductance between a sub-surface cell, i, and a surface cell, j

The formulation of conductance between a sub-surface cell, i, and a surface cell, j is very close to that presented above. The flow resistance in the surface cell is usually insignificant compared to the flow resistance in the sub-surface cell, and is therefore ignored. The conductance between two horizontal cells is given as

$$C_{i,j} = \frac{b_{i,j} dz_{i,j}}{a_{i,j}} K_{h,i} \tag{2.14}$$

and the conductance between two vertical cells are given as

$$C_{i,j} = \frac{A_{poly,i}}{dz_i/2} K_{v,i} \tag{2.15}$$

Conductance between two surface cells

The surface flow is formulated according to diffusive wave theory where the flow, $Q_{i,j}$ is described by Eq. 2.6, where the friction slope, S_f, is approximated with the water level gradient, $(\psi_j - \psi_i)/a_{i,j}$:

$$Q_{i,j} = \text{sign}\left(\frac{\psi_j - \psi_i}{a_{i,j}}\right) M_{i,j} R_{i,j}^{\frac{2}{3}} A_{i,j} \sqrt{\left|\frac{\psi_i - \psi_j}{a_{i,j}}\right|} \tag{2.16}$$

where the Manning number, $M_{i,j}$, the hydraulic radius, $R_{i,j}$ and the wetted cross section area in the cell, $A_{i,j}$ are the arithmetic mean of the cell values of M, R and A.

The calculation of the hydraulic radius and the cross section area depends on whether the cells are river cells or ordinary surface cells. River cells have user-specific cross sections from which the water level dependence of R_j and A_j is calculated. In the ordinary surface cells R_j and A_j are calculated as

$$\begin{aligned} A_j &= b_{i,j}(\psi_j - z_j) \\ R_j &= A_j/P_j = \frac{b_{i,j}(\psi_j - z_j)}{b_{i,j}} = \psi_j - z_j \end{aligned} \tag{2.17}$$

The flow described by Eq. 2.5.1 cannot directly be rewritten in the form of Eq. 2.10 due to the square root of the water level gradient. The problem can be solved either by solving Eq. 2.5.1 iteratively with $\sqrt{|(\psi_i - \psi_j)/a_{i,j}|}$ formulated explicitly (from the last iteration or time step) or by assuming that the water level can be approximated with the river bottom gradient. For reasons of stability

the latter is assumed and Eq. 2.5.1 can be rewritten in the form of Eq. 2.10 if we say that

$$\text{sign}\left(\frac{\psi_i - \psi_j}{a_{i,j}}\right)\sqrt{\frac{|\psi_i - \psi_j|}{a_{i,j}}} \simeq \frac{\frac{\psi_i - \psi_j}{a_{i,j}}}{\sqrt{\frac{|z_i - z_j|}{a_{i,j}}}} \tag{2.18}$$

resulting in the conductance

$$C_{i,j} = \frac{M_{i,j} R_{i,j}^{\frac{2}{3}} A_{i,j}}{a_{i,j}\sqrt{\frac{|z_i - z_j|}{a_{i,j}}}} \tag{2.19}$$

2.5.2 Sink/source terms

The sink/source term, Q_{out}, refers to any exchange between the integrated model and the surroundings. Net precipitation and groundwater abstraction are two examples of such exchanges. The model is capable of having abstractions with filters covering several computational layers. The extracted water is distributed among the layers according to the length of the filter in each layer.

2.5.3 Storage terms

In transient modelling the storage terms are included in the water balance equation, Eq. 2.9. The storage formulation depends on the cell type. For river cells the storage term is

$$\Delta sto_j = (\psi_j^n - \psi_j^{n-1}) A_{river,j} \tag{2.20}$$

where $A_{river,j}$ is the area of the water surface within cell j and will typically vary with the water level. $A_{river,j}$ is found using an average potential between time step $n-1$ and n, $\left((\psi_j^n + \psi_j^{n-1})/2\right)$.

The storage term for ordinary surface cells is given as

$$\Delta sto_j = (\psi_j^n - \psi_j^{n-1}) A_{poly,j} \tag{2.21}$$

and finally the storage term for sub-surface cells is given as (Harbaugh et al. 2000)

$$\Delta sto_j = S2_j(\psi_j^n - z_j^+) - S1_j(\psi_j^{n-1} - z_j^+) \tag{2.22}$$

where $S1_j$ is the storage capacity at the start of the time step and $S2_j$ is the storage capacity at the current iteration.

For combined cells $S1_j$ and $S2_j$ are given as $A_{poly,j}\Delta z S_s$, and for unconfined cells as $A_{poly,j}S_y$. Here S_s and S_y are specific storage and specific yield respectively.

2.5.4 Vertical flow under de-watered conditions

In Eq. 2.15 and Eq. 2.10 the formulation of vertical flow is based on the assumption that the receiving cell is fully saturated and that the hydraulic gradient is $2(\psi_i - \psi_j)/(dz_i + dz_j)$.

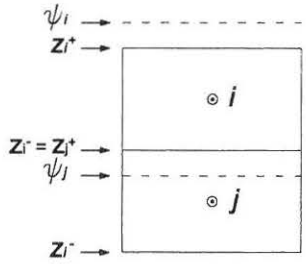


Figure 2.6: De-watering of receiving cell

When the receiving cell is de-watered, Fig. 2.6, the actual flow is

$$Q_{i,j} = (\psi_i - z_i^-)C_{i,j} \tag{2.23}$$

From Eq. 2.10 and Eq. 2.23 the flow is overestimated by

$$Q^*_{i,j} = (\psi_i - \psi_j)C_{i,j} - (\psi_i - z_i^-)C_{i,j} = (z_i^- - \psi_j)C_{i,j} \tag{2.24}$$

The correction term $Q^*_{i,j}$ is added to the sink/source term $Q_{out,j}$

2.5.5 Initial and boundary conditions

The initial conditions are specified potential head in all calculation cells.

At present the no-flow boundary condition (Neumann condition) and the constant head boundary condition (Dirichlet condition) are implemented. If no other conditions are given, the Neumann condition is implicitly implemented at the boundary. The Dirichlet condition is implemented through special head boundary cells.

2.5.6 Solving the finite difference equations

The combination of the water balance equation and the constitutive equations for sub-surface (Darcy's law) and surface flow (Manning formula) results in a set of finite difference equations. The equations are linearised by using the last calculated vector of the system state variables in the calculation of the conductance. The linearised set of equations takes the form

$$\mathbf{A}\psi^n = \mathbf{b} \tag{2.25}$$

where ψ^n is a vector of the unknown system state variables at time step n, $\boldsymbol{A}$ the coefficient matrix and $\boldsymbol{b}$ the constant vector.

Due to the unstructured grid, the coefficient matrix $\boldsymbol{A}$ is sparse and non-diagonal.

The linearised finite difference equations are solved using the NSPCG package, (Oppe et al. 1988), with the Jacobi preconditioner and the conjugate gradient accelerator. The Jacobi preconditioner allows coordinate storage of the coefficient matrix, which is essential for efficiently solving a large equation system with an unstructured coefficient matrix.

The original equations (before linearisation) are highly non-linear and Eq. 2.25 is therefore updated in an iterative procedure.

2.6 Transport modelling

A particle tracking (**PT**) model is implemented in FLEXFLOW for the purpose of determining capture zones.

2.6.1 Governing equation

The governing equation for the advection and dispersion process is given as (Note that as an exception, index notation is used in the description, $x_1 = x$, $x_2 = y$, $x_3 = x$, $D_{1,2} = D_{x,y}$, etc.)

$$\frac{\partial c}{\partial t} + v_i \frac{\partial c}{\partial x_i} + \frac{\partial}{\partial x_i}\left(D_{i,j}\frac{\partial c}{\partial x_j}\right) = \frac{G}{\theta_{eff}} \tag{2.26}$$

where

v_i	water velocity,[m s^{-1}], in i direction, where $i = (x, y, z)$. $v = Q/A$ for surface flow and $v = q/\theta_{eff}$ for sub-surface flow. A is the cross section area and q is Darcy velocity (filter velocity)
c	concentration [kg m^{-3}]
$D_{i,j}$	components in the hydrodynamic dispersion matrix [m^2 s^{-1}]. $D_{i,j}$ is the sum of the mechanical dispersion and the molecular diffusion coefficient.
G	sink/source term [kg m^{-3} s^{-1})]
θ_{eff}	effective porosity [m^3 m^{-3}]. For surface flow, θ is one or close to one, and for sub-surface flow, θ is the volume of mobile water (mobility here is in relation to the pressure gradient that is present in sub-surface flow).

2.6.2 Particle tracking method

Typically the advection-dispersion equation (2.26) is solved by finite difference, finite element or particle tracking (random walk) techniques. Great care is necessary in order to avoid numerical errors and numerical dispersion in relation to finite difference and finite element modelling. Often the grid resolution has to be very fine in order to achieve accurate results. As its name indicates, the particle tracking method is a method based on particles to represent the solute and the displacement of particles by the transport processes of advection and dispersion. The method is grid independent and the accuracy depends on the number of particles transported.

The particle tracking method is widely used in groundwater modelling aimed at determining capture zones.

In FLEXFLOW a particle tracking method for the coupled surface and sub-surface flow has been implemented.

Displacement of particles

An essential step in the PT method is the movement of the single particle so that the swarm of particles represents the solution of the advection dispersion equation. The displacement vector, $\Delta\mathbf{x} = (\Delta x, \Delta y, \Delta z)^T$ at a given point, (x_0, y_0, z_0),

is (see e.g. Prickett et al. (1981)):

$$\begin{aligned}
\Delta x = &\ v_x \Delta t + r_1 \sqrt{2\alpha_{x,x} v_x \Delta t} + r_2 \sqrt{2\alpha_{x,y} v_y \Delta t} + r_3 \sqrt{2\alpha_{x,z} \Delta t} \\
\Delta y = &\ v_y \Delta t + r_1 \sqrt{2\alpha_{y,x} v_x \Delta t} + r_2 \sqrt{2\alpha_{y,y} v_y \Delta t} + r_3 \sqrt{2\alpha_{y,z} \Delta t} \\
\Delta z = &\ v_z \Delta t + r_1 \sqrt{2\alpha_{z,x} v_x \Delta t} + r_2 \sqrt{2\alpha_{z,y} v_y \Delta t} + r_3 \sqrt{2\alpha_{z,z} \Delta t}
\end{aligned} \tag{2.27}$$

where

r_1, r_2,r_2	random unit Gaussian distributed numbers [-]
v_x, v_y, v_z	velocities [m/s] (pore velocities in sub-surface flow)
$\alpha_{.,.}$	dispersivity [m]
Δt	time step [s]

In order to calculate the particle displacement at an arbitrary point the water velocity at that point has to be known. The flow at the cell interfaces is known from the Darcy equation and the transformation into water velocities at an arbitrary point is divided into two steps. Step one is to determine the x and y velocity components at the cell node, and step two is to interpolate the node velocities at the desired point at which flow velocities are required.

Step 1 - node velocities

Let us first consider the the x-component of the node velocity. In the plan view the cell is subdivided into two by an imaginary line, I_y, through the node, Fig. 2.7a. Water balance considerations are used to find the approximate flow, $Q_{x,j}$ through I_y.

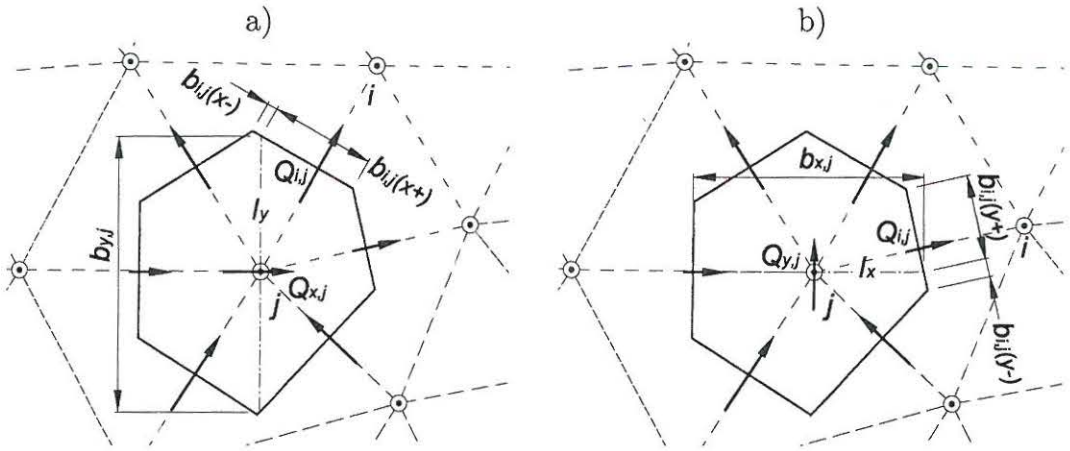


Figure 2.7: Definition sketch for calculation of nodal flow velocity. a) flow velocity in x-direction, b) flow velocity in y-direction

The water balance for the left sub-region is

$$Q^{-}_{x,j} - Q_{x,j} = -\frac{Q_{out,j}}{2} + \frac{\Delta sto_j}{2\Delta t} \tag{2.28}$$

and for the right sub-region

$$Q_{x,j} - Q^{+}_{x,j} = -\frac{Q_{out,j}}{2} + \frac{\Delta sto_j}{2\Delta t} \tag{2.29}$$

Combining Eq. 2.28 and 2.29 yields

$$Q_{x,j} = \frac{Q^{+}_{x,j} + Q^{-}_{x,j}}{2} \tag{2.30}$$

where $Q^{-}_{x,j}$ is the boundary net inflow in the left sub-region (horizontal and vertical contributions) and $Q^{+}_{x,j}$ is the boundary net outflow in the right sub-region (horizontal and vertical contributions). When referring to the known Darcian boundary flow $Q^{-}_{x,j}$ and $Q^{+}_{x,j}$ become

$$\begin{aligned} Q^{-}_{x,j} &= \sum_{i=1}^{N_{cell,j}} Q_{i,j} \frac{A^{x-}_{i,j}}{A_{i,j}} \\ Q^{+}_{x,j} &= \sum_{i=1}^{N_{cell,j}} Q_{i,j} \frac{A^{x+}_{i,j}}{A_{i,j}} \end{aligned} \tag{2.31}$$

where

$A_{i,j}$	area of interface between cells i and j (flow section area)
$A^{x-}_{i,j}$	area of interface between cells i and j that is placed to the left of line l_x. For vertical flow $A^{x-}_{i,j} = A_{poly,j}/2$ and for horizontal flow $A^{x-}_{i,j} = b_{i,j}(x-)dz_{i,j}$
$A^{x+}_{i,j}$	area of interface between cells i and j that is placed to the right of line l_x. For vertical flow $A^{x+}_{i,j} = A_{poly,j}/2$ and for horizontal flow $A^{x+}_{i,j} = b_{i,j}(x+)dz_{i,j}$
$b_{i,j}(x-)$	fraction of $b_{i,j}$ left of I_y, Fig. 2.7a
$b_{i,j}(x+)$	fraction of $b_{i,j}$ right of I_y
$N_{cell,j}$	number of cells interacting with cell j

The water velocity, $v_{x,j}$, at node j is

$$v_{x,j} = \frac{Q_{x,j}}{2b_{y,j}\min(dz_j, \psi_j - z^{-}_j)\theta_{eff,j}} \tag{2.32}$$

where

$b_{y,j}$	cell height at nodal point of cell j
ψ_j	head potential or water level in cell j
$\theta_{eff,j}$	effective porosity.

Combining Eq. 2.30 and 2.32 yields

$$v_{x,j} = \left(\sum_{i=1}^{N_{cell,j}} Q_{i,j} \frac{A_{i,j}^{x-}}{A_{i,j}} + \sum_{i=1}^{N_{cell,j}} Q_{i,j} \frac{A_{i,j}^{x+}}{A_{i,j}} \right) \frac{1}{2b_{y,j} \min(dz_j, \psi_j - z_j^-)\theta_{eff,j}} \tag{2.33}$$

and a similar expression for the y-component can be found, (Fig 2.7b) as

$$v_{y,j} = \left(\sum_{i=1}^{N_{cell,j}} Q_{i,j} \frac{A_{i,j}^{y-}}{A_{i,j}} + \sum_{i=1}^{N_{cell,j}} Q_{i,j} \frac{A_{i,j}^{y+}}{A_{i,j}} \right) \frac{1}{2b_{x,j} \min(dz_j, \psi_j - z_j^-)\theta_{eff,j}} \tag{2.34}$$

Step 2 - interpolation

The x and y velocities at point (x_0, y_0, z_0) are found by bi-linear interpolation between the velocities in the three nodes that constitute the triangle in which the point is located, Fig. 2.8 and Eq. 2.35.

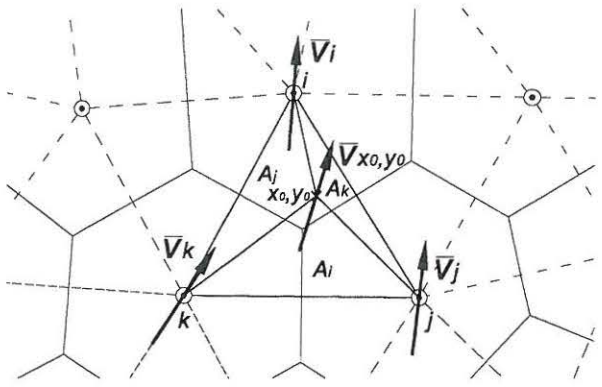


Figure 2.8: Bi-linear interpolation of horizontal flow velocities

$$\begin{aligned} v_x(x_0, y_0) &= \tfrac{1}{A_i + A_j + A_k}(A_i v_{x,i} + A_j v_{x,j} + A_k v_{x,k}) \\ v_y(x_0, y_0) &= \tfrac{1}{A_i + A_j + A_k}(A_i v_{y,i} + A_j v_{y,j} + A_k v_{y,k}) \end{aligned} \tag{2.35}$$

where A_i, A_j and A_k are the areas of the triangles opposite nodes i, j and k, respectively, see Fig. 2.8.

The vertical flow velocity at point (x_0, y_0, z_0) is determined by linear interpolation of the vertical velocities at the cell interfaces

$$v_z(x_0, y_0, z_0) = \frac{1}{dz_j}((z_0 - z_j^-)v_z^+ + (1 - (z_0 - z_j^-))v_z^- \tag{2.36}$$

where v_z^- is the flow velocity at the bottom of cell j, and v_z^+ is the flow velocity at the top of cell j, both being positive upwards. See Fig. 2.5b for definition of z_j^-, z_j^+ and dz_j.

2.6.3 Verification of flow and transport models

FLEXFLOW has been tested against the Theis solution (see e.g. Bear (1979)) and the solution to the draw down in relation to abstraction in the centre of a circular island under confined conditions. A high level of agreement between simulated results and analytic solutions was found.

The particle tracking module has so far been used only for simulating advective flow transport. The module has been tested in a uniform, stationary flow domain in order to verify the advective velocity of the particles. The module has furthermore been tested in a 3-dimensional heterogeneous aquifer, see Chapter 5 for setup. Particles were distributed uniformly on the surface and the number of particles leaving the flow domain respectively as abstracted water, through the river, and over the boundary, was registered. Every particle represents an amount of infiltrated water, and thus the relative number of particles leaving the domain, e.g. as abstracted water, must be similar to the relative amount of water leaving the domain through the abstraction, or in other words, Eq. 2.37 has to be fulfilled.

$$\frac{N_{part,abs}}{N_{part}} \simeq \frac{Q_{abs}}{Q_p} \tag{2.37}$$

where $N_{part,abs}$ is the number of particles abstracted, N_{part} is the total number particles added to the flow domain, Q_{abs} is the amount of abstracted water and Q_p is the amount of precipitation. A high level of agreement was found in both tests.

CHAPTER 3

Classical calibration procedures

I argued in the introduction that calibration and uncertainty estimation are closely connected. Conceptual model construction and parameterisation are usually founded on a weak data base. Values derived from literature and from personal experience may narrow the range of possible conceptual models and values of parameters and variables. Nevertheless, the variety of solutions may still be too large to provide acceptable predictions.

Calibration is the process of rating different alternative sets of conceptual models, parameters and variables according to the degree of fit between simulation and observations. Classical calibration procedures aim to find a single unique parameter set that corresponds to the global optimum with regard to the degree of fit.

The parameter estimating problem has traditionally been solved by trial-and-error techniques, with the hydrologist successively changing the unknowns until he/she believes that the solution is close to the optimal. For large complex models with many unknowns this trial-and-error calibration procedure is very difficult and time consuming. With the constant increase in computer power, a range of automatical calibration techniques becomes more and more interesting.

This chapter gives an introduction to objective functions within standard calibration procedures as well as to the statistics of calibrated parameters and predicted state variables. The issue of single versus multi-objective function is briefly discussed.

First, however, I look at the issue of parameterisation and the different types of observation data that can be used. The concept of ill-posedness is defined and

model non-linearity is discussed.

3.1 Parameterisation

In distributed physically-based groundwater models the domain under consideration is discretised into a finite number of elements/cells, each of which has an individual set of parameters - i.e. sub-surface cells have hydraulic conductivities in three directions, a storage coefficient, effective porosity, sink/source term, etc.

Parameterisation concerns the assignment of values to these parameters. In most model applications the scale of parameterisation is larger than the scale of the numerical cells, i.e. parameters are assumed to be constant within regions/zones that are larger than the cell scale. In relation to calibration and uncertainty estimation the issue of parameterisation is very important.

3.1.1 The continuum approach

The continuum approach is fundamental in groundwater modelling and introduces the first set of restrictions at the level of parameterisation. The flow process on the molecular/microscopic scale, Fig. 3.1, is very complicated and in practice inaccessible. First, it is impossible to determine the exact geometry of the soil under consideration and, secondly, it is impossible to solve the flow equations for the volumes normally considered in groundwater modelling problems. Using a cylindrical pipe for stationary flow, Henry Darcy found by experiment a linear relation between the head gradient and the flow.

$$q = \frac{Q}{A} = -\frac{\Delta\psi}{\Delta L}K \tag{3.1}$$

where q is the Darcy or filter velocity, $\Delta\psi$ is the head difference over the soil sample, ΔL is the length of the soil sample, A is the cross section area of the soil sample (similar to the pipe cross section area) and K is the proportional constant or hydraulic conductivity.

Darcy's law is based on a macroscopic scale and builds on the continuum approach, where the porous media are considered as a continuum for which representative average parameters can be found. By introducing Darcy's law we are prevented from describing flow on anything smaller than the macroscopic scale. For many problems the macroscopic scale is sufficient, but when it comes to considering transport processes the flow on the microscopic scale cannot be ignored because of the high degree of variation in the size of pore velocities. In order to compensate for the unknown variation of flow velocities and directions

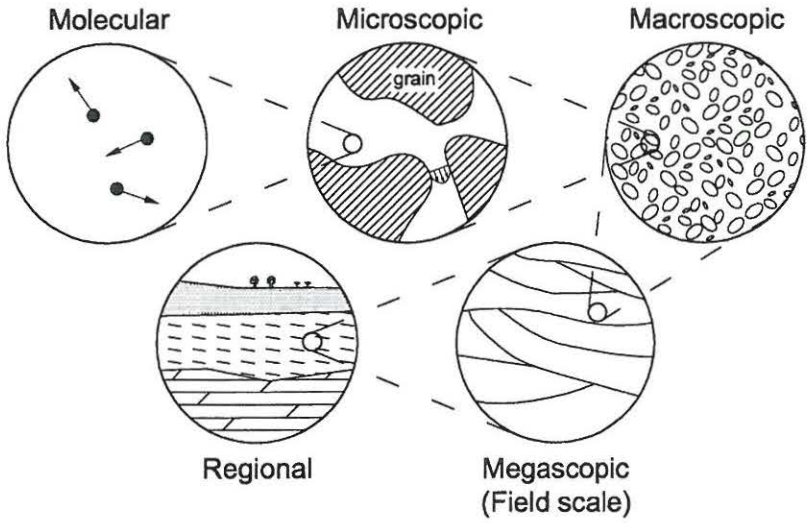


Figure 3.1: Definition of scales.

on the microscopic scale it is essential when modelling transport processes to introduce a dispersion process.

3.1.2 Parameter scale

Darcy's law is the foundation for describing laminar flow in porous media. It is used directly in the derivation of the governing equations for groundwater flow, and in the laboratory for estimating the flow properties of a given soil sample. Darcy's law combined with a mass balance equation, together with the relevant dispersion relations, constitute the governing equations on a macroscopic scale. It is thus straightforward to discretise and solve the equations with related parameters on this scale. However, in most cases the discretisation scale is much coarser than the macroscopic scale, and there is no guarantee that parameter values on the measured scale are representative of the model scale. The term "effective parameter" is often used for parameters on the model scale, to suggest the fact that the parameters have no direct physical meaning. Figure 3.2 illustrates the typical evolution in a given parameter as a function of averaging volume. It can be seen that a number of different Representative Elementary Volumes (R.E.V.) can be found, depending on the scale under consideration.

In principle measured parameters can be used in the governing equations only on the scale on which they are found. So the question is: is the scale on which parameters are measured identical with the scale of the numerical models applied to a given problem? There is no general answer to this question, but in most cases the answer is negative.

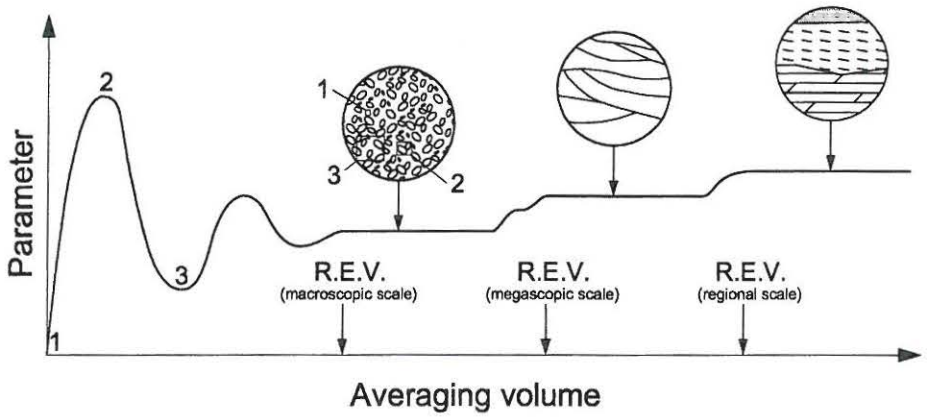


Figure 3.2: Parameter value versus averaging volume.

Figure 3.3 is an attempt to illustrate the parameter scale in relation to the method of measurement and the scale of application, which depends on which type of model is used.

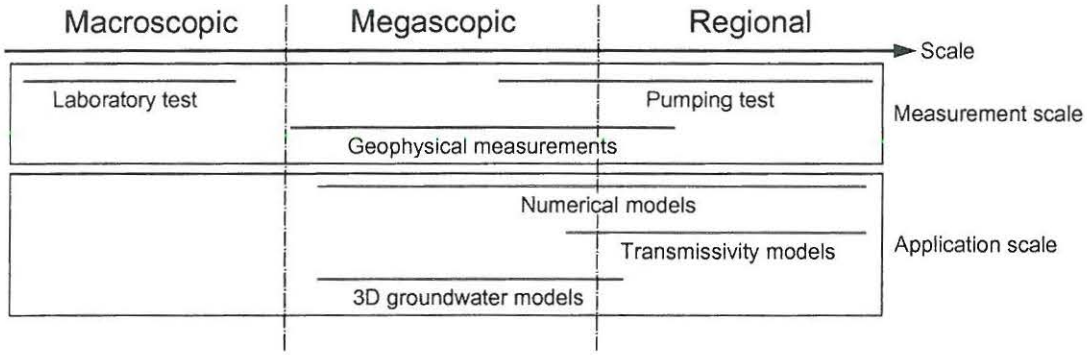


Figure 3.3: Measurement and application scales.

Geophysical measurements are in general used only to identify geological units, and it is very difficult to use them as the basis for estimating hydrogeological parameters. So long as interpretative assumptions are fulfilled, pumping tests are cabable of predicting aquifer properties on the megascopic and regional scales. Laboratory tests are typically performed on the macroscopic scale, and in the case of undisturbed soil samples the test might give a good estimate of parameter properties on the given scale.

From Figure 3.3 it can be seen that only geophysical measurements seem to have the scale on which 3D groundwater models are normally applied. Given the above mentioned limitations of geophysical measurements, no method of measurement is suited to giving precise parameter estimates on the megascopic scale on which 3D groundwater models are normally implemented.

Thus on the one hand the various methods of measurement may not yield precise parameter estimates while, on the other, the results may substantially reduce the possible parameter range.

3.2 Observation data

Calibration is performed with respect to past system behaviour. System behaviour is typically described by observations in time and space e.g. of water level, head potential and river discharge. But in principle all observations that in some way condition the parameter estimates are valuable. This section describes the different types of observations that can be incorporated into a calibration procedure and, where possible, the sources of mismatch between observed and computed values are described.

3.2.1 Head data

Head and water level observations originate from at least three types of surveys: a) logging in connection to newly established wells, b) more or less continuous logging of existing wells and c) synchronous observations of a large number of wells.

A mismatch between calculated and observed head data may originate from at least three sources: 1) observation errors, 2) scale errors and 3) errors due to time effects. The following description is inspired by Sonnenborg (2001).

Observations errors: The head observation errors in this description consist of errors directly associated with measurement, rather than all sources of mismatch between computed and observed heads, such as may be found in other literature. The contributions to head observation errors are:

- Finite precision of the measuring equipment. The expected error contribution is in the range of a **few millimetres** to a **few centimetres**.
- Manual reading of instruments and registration of the results. The expected error contribution is in the range of a **few millimetres** to a **few centimetres**.
- Incorrect or imprecise well reference level. References levels may be determined with considerable accuracy by using a levelling instrument or GPS. In this case the error contribution is typically in the range of a **few millimetres** to a **few centimetres**. Alternatively the reference level may be

determined from topographical maps and the error contribution may be several **metres**, depending of the quality and resolution of the maps.

- Atmospheric pressure. Head potential in confined aquifers fluctuates in proportion to fluctuations in the atmospheric pressure. The atmospheric pressure fluctuations are rarely taken into account in groundwater models and may lead to an error contribution of up to 0.1 m. The change in potential head due to a change in atmospheric pressure can be estimated from

$$\psi - \psi_0 = -\frac{BE}{\gamma}(p_a - p_{a0}) \tag{3.2}$$

where γ [kg m s^{-2} m^{-3}] is the gravitational body force, BE is the barometric efficiency defined as the ratio of water level change, $(\psi - \psi_0)$ [m], to the atmospheric pressure, $(p_a - p_{a0})$ [bar]. BE has been observed in the range of 0.25 to 0.75. (Bear 1972).

A pressure difference of 20 mb (the passage of a storm depression) and a barometric efficiency of 0.5 yield a water level change of $\sim$ 0.1 m.

Scale errors: Scale errors originate partly from the finite discretisation of the computational cells and partly from the discretisation of the parameters involved. The main contributions to scale errors are:

- Horizontal discretisation. Head values are calculated at the centre of each computational cell and the observed heads may be located anywhere in the computational cell. A comparison of observed heads with the nearest calculated head value may result in an error contribution up to $0.5Jdx$, where J is the horizontal gradient of the hydraulic head and dx is the size of the computational cell. If the computed head is interpolated into the location of the observed head the error contribution is significantly smaller.

- Vertical discretisation. Observed head values are representative of head values ranging from the minimum to the maximum head in the filter section. Typically the head in the high yield formations in a filter section will have a strong influence on what is measured. Simulated heads are average head in the computational layer. Given that the filter section is identical with the computational layer, the maximum error contribution is of the magnitude $0.5J_v dz$, where J_v is the vertical hydraulic gradient over the filter section, and dz is the computational layer thickness. Filter information may in many cases be incomplete and discussions as to which geological unit has actually been measured may occur. The error contribution arising from an imperfect match between filter levels and computational layers is difficult to quantify, since information on vertical gradients is rare. The

error contribution from monitoring the wrong geological unit (e.g. monitoring the secondary aquifer believing that is is the primary aquifer) may be up to **several metres**.

- Topographical variations. Topographical variations within a computational cell may result in variations in the potential head that will not be represented by the numerical model. The error contributions may be significant in near-surface reservoirs, and can be assumed to be proportional to the topographical variations, and inversely proportional to the depth below ground surface (Sonnenborg 2001).

- Hydrogeological heterogeneity. Parameter heterogeneity cannot be resolved at a finer scale than the resolution of the numerical grid. Usually heterogeneity within a geological unit is ignored and the parameters are constants within regions much larger than the grid. The ignored heterogeneity will contribute to the error between computed and observed heads where the standard error, s_h, may be formulated as a function of the hydraulic gradient J, the standard deviation of the log-transformed (natural) hydrological conductivity, $\sigma_{\ln K}$, and the length scale, α_l. The error contribution may be approximated as (Gelhar 1986).

$$s_h = \sqrt{\frac{1}{3} J^2 \sigma^2_{\ln K} \alpha_l^2} \tag{3.3}$$

  The length scale, α should be chosen as the minimum of: a) the correlation length of the conductivity field, b) extension of the region with constant parameters. In the case of fully distributed parameters (with a different value in each cell) the grid dimension should be used.

  Errors due to hydrogeological heterogeneity may alternatively be approximated from head observation in closely positioned wells. The difference in observed heads, subtracting all other sources of error, may represent the errors due to hydrogeological heterogeneity (Sonnenborg 2001).

Errors due to time effects: The number of head observations in time and space is often limited, and all available observations have to be used. Typically the set of chosen head observations is a mixture of equidistant time series, non-equidistant time series (few observations) and single observations. The data set may be incomplete in a variety of ways, and this may contribute to error. The description below focuses on the error contribution in relation to stationary models.

In stationary models time-average head conditions are simulated, and present head observations may not represent stationary conditions. This may lead to errors. The size of these errors can be approximated by analysing the time series for neighbouring wells. The aim of this analysis is to establish yearly fluctuations

and trends, seasonal fluctuations and more rapid fluctuations. Yearly fluctuations and trends represent errors due to adapting head data from one year in the calibration of another year. Seasonal fluctuations represent the errors arising from presenting a seasonal representative observation as the average yearly head. The rapid fluctuations represent errors due to ignored small time-scale variations e.g. in infiltration and abstraction. Useful statistics could be: a) average weekly standard deviation of daily values, b) average seasonal standard deviation of weekly averages, c) average yearly standard deviation of seasonal averages, d) trends in yearly average. Sonnenborg (2001) has suggested $\Delta H/2$, where ΔH is the difference between the minimum and maximum value in the time series, as a simple measure of the error contribution.

Table 3.1 summarises the different contributions to the mismatch between observed and computer head. Values and intervals are approximate and based on Danish conditions.

Table 3.1: Standard deviation of error contributions. Inspired by Sonnenborg (2001)

	Error contribution	Standard deviation [m]	Reference
Observation error	Measuring equipment	0.03	
	Reading and bookkeeping	0.05	
	Reference level	0.05 - 2.0	
	Atmospheric pressure	0.0 - 0.15	
Scale error	Horizontal discretisation	$0.5\Delta x J$ [1)]	Sonnenborg 2001
	Vertical discretisation	$0.5 J_v dz - 2.0$ [2)]	
	Topographical variations	σ_{topo}/d [3)]	Sonnenborg 2001
	Heterogeneity	$\sqrt{\frac{1}{3} J^2 \sigma^2_{lnK} \alpha_l^2}$ [4)]	Gelhar 1986
Time effects	Non-stationarity	$\Delta H/2$ [5)]	Sonnenborg 2001
Total error		$\sqrt{\sum \sigma^2}$	

1. Δx is the horizontal discretisation and J is the hydraulic gradient.
2. Δy is the vertical discretisation and J_v is the vertical gradient.
3. σ_{topo} is the standard deviation of the topography within a computational grid. d is the depth from ground surface to groundwater.
4. $\sigma_{\ln K}$ is the standard deviation of the log-transformed (natural) hydrological conductivity. α is the minimum of the correlation length for $\ln K$ and extension of the region with constant parameters.
5. ΔH is the difference between the minimum and maximum value in the time series.

3.2.2 River discharges

Measurements of river discharge are regularly carried out in major streams and rivers. This is done by measuring water levels and deriving the actual flow from these measurements, using a mathematical relation between water level and flow (Q-ψ relation). These point flow velocity measurements are then transformed into discharges. The observation data originate from a number of continuously logged gauging stations and/or from synchronous measurements of the flow in a large number of cross sections.

As in the case of head data, the mismatch between calculated and observed river discharges may originate from observation errors, scale errors and errors due to time effects.

Observation errors: Observation errors originate from the registration of water levels, the registration of flow velocities and the subsequent transformation of these measurements into discharges.

- Discharge derived from water level registration. Water level measurements are transformed into discharges by using a Q-ψ relation, describing the relationship between water level and discharge. The Q-ψ relation for a given cross section is dynamic and will change with the amount of vegetation and possible erosion. The uncertainty of the discharge estimate is in the order of 10% (Blicher 1991).

- Discharge derived from flow velocity measurements. When discharge is estimated from a number of point velocity measurements in well defined cross sections, the level of uncertainty is relatively low, probably in the order of 5%.

Scale errors: Scale errors in relation to river modelling are closely related to numerical and parameter discretisation.

- The level of detail. The level of detail at which the geometry of a river can be described depends on numerical discritisation. River branches smaller than grid dimension cannot be described.

- Topographical variations. The interaction between river and groundwater may depend to a considerable degree on small-scale variations in the topography: e.g. seepage flow in local topographical depressions.

- Hydrogeological heterogeneity. Small-scale heterogeneity may be determined for the interaction between groundwater and river.

Definitions

Median value of annual minimum flow is the daily median value of the annual minimum flow.

Error due to time effects: Synchronous observations of summer discharge are performed for the purpose of estimating the base flow contribution to the river. The results may be converted to another year or years from a reference gauging point, on the assumption that the synchronous observation points behave in the same manner as the reference station. Errors may occur, depending on the degree to which this assumption is fulfilled. (Sonnenborg 2001)

Time series data are used to estimate the median value of annual minimum flow. Bjarnov (1987) has established the relation between base flow, q_b $[l/s/km^2]$, and the standard deviation on the median value of annual minimum flow, s_{q_b} $[l/s/km^2]$ reported in Fig. 3.4, for ten Danish stationary gauging stations.

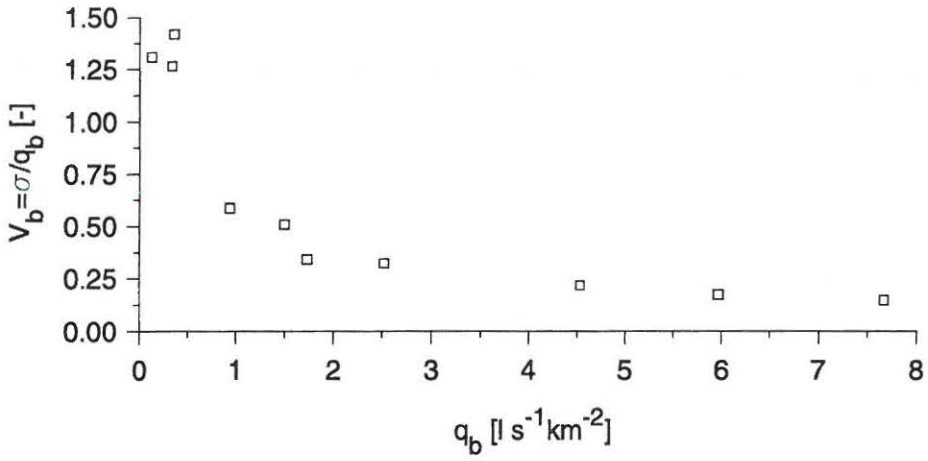


Figure 3.4: Relation between the median value of annual minimum flow, q_b, and the coefficient of variation of the median value of annual minimum flow, s_q for 10 Danish stationary gauging stations. Estimation based on a 65-year time series

3.2.3 Concentration data

Concentration data are available from regular testing in abstraction wells and from surveys of polluted sites. Concentration data can be valuable as tracer data in cases where the source and leakage period are known, for example in the case of gasoline pollutions with MTBE spill. MTBE was added to gasoline in the mid-eighties and the source location is often well described. The calibration of concentration levels from point pollutions is in general associated with large

errors due to: 1) the density of the concentration data in time and space in comparison with the extension of the pollution, 2) great uncertainty in the description of degradation and sorption, 3) hydrological heterogeneity. The effects of 1) and 3) may be averaged out if area sources, such as nitrate or pesticides, are considered.

3.2.4 Tracer data (natural and artificial)

Natural tracer data such as CFC gas may be very useful in determining groundwater age. The uncertainty of the age estimate depends on the degree of mixture. The water in abstraction wells is typically a mixture of water from large regions with different travel time, and for this reason estimates of its age may be very uncertain. In observation wells with narrow filter intervals one can assume a lower level of uncertainty, since the only mixture to occur here is natural.

Because groundwater motion is usually slow, artificial tracers are often used over short distances, normally in order to constrain the flow velocities in a given deposit. The uncertainty of flow velocity estimates is due to at least two factors: the finite number of observation points and the existence of small-scale hydrological heterogeneity.

3.2.5 Subjective observations

Observations by local farmers and citizens offer an alternative form of data where hydrological conditions are otherwise ungauged. A local farmer may for example observe that the "creek on his land dries out every summer" or that his "spring yields water only in wet years". It is challenging to incorporate subjective observations into an automatic calibration procedure, but such information may be very valuable.

3.2.6 Weighting of observations

If the aim in calibration is to incorporate different kinds of observations (e.g. head and river discharge) and/or observations of the same kind with varying levels of expected error, the observations have to be mutually weighted. The standard procedure is to weight the observations according to the inverse of the estimated error covariance matrix, $\mathbf{V} = \mathbf{C}^{-1}$.

The correlation between observations is usually very difficult to determine and in general observations are assumed to be uncorrelated. Positive correlation may, however, be a useful tool in ensuring certain trends. In many situations, for

example, it will be equally crucial to ensure a low average mean value of the absolute head residuals and at the same time a high level of agreement between the observed head gradients and simulated head gradients. A high degree of positive correlation will ensure weighting of the gradients.

3.3 Ill-posedness

Well-posedness is a fundamental requirement in applying traditional calibration procedures to a calibration problem. A well-posed calibration problem is defined as one that has a parameter solution that is **identifiable**, **unique** and **stable**.

The solution is identifiable if it can be found from the observation set, and it is unique if only one such solution can be found.

The solution is thus unique if one and only one set of models, parameters and variables can be established from the set of observations. A necessary, but not sufficient, requirement for uniqueness is that the number of models, parameters and variables to be estimated is less or equal to the number of observations. The observations have to be spatially distributed and preferably of different types (e.g. head, stream flow, tracer, etc.).

The solution is stable if small changes in the observations produces small changes in the parameter solutions. Instability often arises from the fact that the parameter solution is non-identifiable, or only poorly so. It is generally associated with objective functions that are flat or nearly flat in the region around the parameter optimum. Unstable problems can result in parameter solutions that are very sensitive to the starting point of the search (Carrera and Neuman 1986b).

A thorough description of (in)stability, (non-)uniqueness and (non-)identifiability can be found in Carrera and Neuman (1986b) and Yeh (1986).

As the above indicates, ill-posedness thus depends on parameterisation - any model can be rendered well-posed by reducing the number of parameters. One should however bear Fig. 3.5 in mind.

Figure 3.5 shows the contributions to the total modelling error as a function of the number of parameters involved. As the number of parameters increases the error contribution from model error decreases, and the contribution from parameter error increases. When the model in simplified in order to make the calibration problem unique, there is a corresponding increase in the contribution from model error.

Example 3.1 *If we consider example 1.1 on page 7 again, and assume that only*

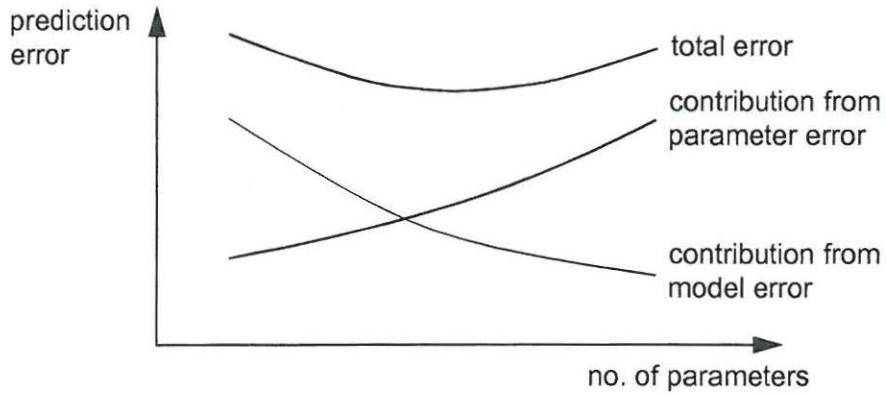


Figure 3.5: Contributions to the total modelling error as a function of the number of parameters

head observations are present, this leaves us with two observations, h_2^* and h_3^*, and two unknowns, q and T. The first requirement of uniqueness is fulfilled (no. of observations is greater than or equal to the no. of unknowns). However, the problem is non-unique because only one type of observation (head) is present. An infinite number of combinations of q and T will result in the same head distribution - only the ratio between q and T can be found. The problem becomes unique if the river inflow estimate, Q_r, is included. Alternatively we could consider q as being deterministic and take only T as the target of calibration. By assuming that q is known, we increase the amount of model error.

3.4 Non-linearity and discontinuity

Groundwater problems are in general non-linear. This non-linearity results from: 1) changing water levels in free reservoirs, leading to changes in cross section area/transmissivity, 2) discontinuous or non-linear sink-source terms, 3) spatial variations in the hydrogeological parameters in the flow domain and 4) threshold-dependent processes such as groundwater pumping, drainage flow or surface flow. Figure 3.6 illustrates a schematic aquifer system and the response of the head potential in the lower aquifer as a function of the change of the hydrological conductivity in the lower aquifer.

At low conductivities only a small amount of water will leave the model through the lower aquifer. The model will instead generate surface flow, drain flow and horizontal flow in the upper aquifer. As the conductivity increases the model predicts a drop in the head potential in the lower aquifer and for large conductivities horizontal flow will only appear in the lower aquifer. In some conductivity regions the head response (response surface) may be close to linear, while in oth-

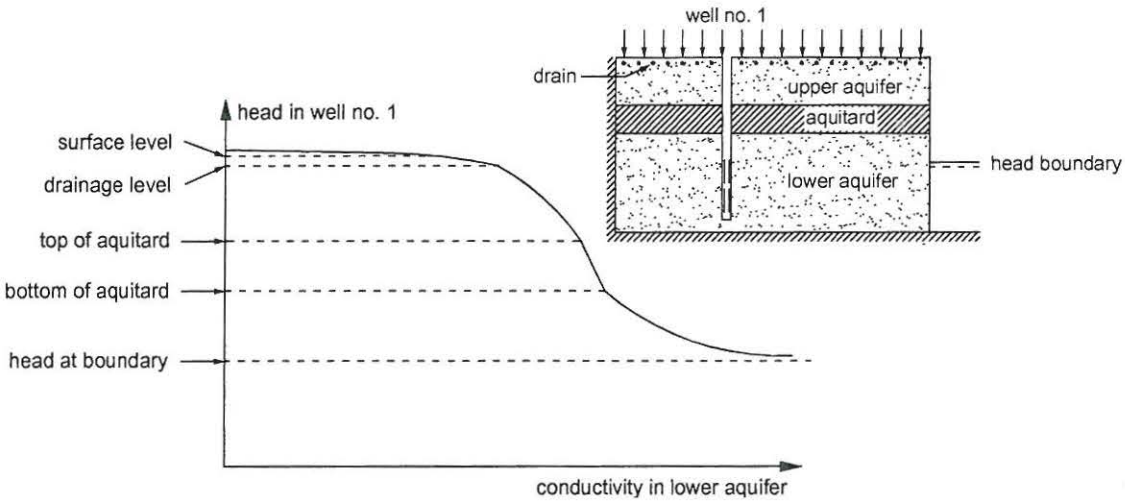


Figure 3.6: Example of head development in an aquifer due to changes in the conductivity - effect of threshold-dependent processes, discontinuous sink-source terms, spatially varied hydrogeological parameters and transmissivity.

ers it will be strongly non-linear. A plateau and a valley in the response surface are found at respectively low and high conductivity values.

This imaginary example illustrates a response surface that has a plateau, a valley and varying degree of non-linearity. Furthermore the response surface is non-differentiable in a number of points. These circumstances play an important role in estimating parameters. Plateaus, valleys and a rapid change in the gradient of the response surface may result in problems when using gradient search methods for optimisation, and models are often linearised in order to stabilise the optimisation: see e.g. Christensen et al. (1998). The subsequent estimation of parameter statistics and the prediction of uncertainties are often based on linearised estimates, which may be misleading for strong non-linear models.

3.5 Least square method

Non-linear least square methods are widely used for parameter estimation in groundwater models (Cooley 1977; Cooley 1979; Hill 1992). The standard least square parameter estimate results from a minimisation of

$$J(\boldsymbol{\theta}) = [\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})]^T \mathbf{C}_{\boldsymbol{\psi}^*}^{-1} [\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})] \tag{3.4}$$

where $\boldsymbol{\psi}^*$ is the vector of observed state variables, $\boldsymbol{\psi}(\boldsymbol{\theta})$ is the vector of computed state variables given the parameter set $\boldsymbol{\theta}$, and $\mathbf{C}_{\boldsymbol{\psi}^*}$ is the expected error covariance of the observed state variables.

If any prior parameter information exists Eq. 3.4 is expanded to

$$J(\boldsymbol{\theta}) = (\mathbf{z}^* - \mathbf{z})^T \mathbf{C}_{z^*}^{-1} (\mathbf{z}^* - \mathbf{z}) \tag{3.5}$$

where

$$\begin{aligned} \mathbf{z}^* &= (\boldsymbol{\psi}^*, \boldsymbol{\theta}^*)^T \\ \mathbf{z} &= (\boldsymbol{\psi}, \hat{\boldsymbol{\theta}})^T \\ \mathbf{C}_{z^*} &= \begin{pmatrix} \mathbf{C}_{\boldsymbol{\psi}^*} & 0 \\ 0 & \mathbf{C}_{\boldsymbol{\theta}^*} \end{pmatrix} \end{aligned}$$

$\boldsymbol{\theta}^*$ is the expected value of the prior parameter vector, $\mathbf{C}_{\boldsymbol{\theta}^*}$ is the covariance of the prior parameter estimate and $\hat{\boldsymbol{\theta}}$ the parameter estimate.

Example 3.2 *In example 1.1, page 7, we want to estimate q and T from the observations ψ_2^*, ψ_3^* and Q_r^* including the prior information on q^* and T^*. From the various sources of uncertainty described in section 3.2 the expected error of ψ_2^*, ψ_3^* and Q_r^* is estimated to $\sigma_{\psi_2^*}$, $\sigma_{\psi_3^*}$ and $\sigma_{Q_r^*}$. From example 1.1 we have the prior information on q^* and T^* $(\mu_q^*, \sigma_q^*, \mu_T^*, \sigma_T^*)$. The transmissivity T is log-transformed before estimation. If the observations and parameters are assumed to be uncorrelated, the components of Eq. 3.5 are given as*

$$\begin{aligned} \mathbf{z}^* &= (\psi_2^*, \psi_3^*, Q_r^*, \mu_q, \mu_{\ln T})^T \\ \mathbf{z} &= (\psi_2(\hat{q}, \hat{T}), \psi_3(\hat{q}, \hat{T}), Q_r(\hat{q}, \hat{T}), \hat{q}, \hat{T})^T \\ \mathbf{C}_{z^*} &= \begin{pmatrix} \sigma_{\psi_2^*} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\psi_3^*} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{Q_r^*} & 0 & 0 \\ 0 & 0 & 0 & \sigma_q & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\ln T} \end{pmatrix} \end{aligned}$$

when inserted in Eq. 3.5 this gives

$$\begin{aligned} J(\hat{q}, \hat{T}) =& \frac{1}{\sigma_{\psi_2^*}} (\psi_2^* - \psi_2(\hat{q}, \hat{T}))^2 + \frac{1}{\sigma_{\psi_3^*}} (\psi_3^* - \psi_3(\hat{q}, \hat{T}))^2 + \frac{1}{\sigma_{Q_r^*}} (Q_r^* - Q_r(\hat{q}, \hat{T}))^2 \\ &+ \frac{1}{\sigma_g} (\mu_q - \hat{q})^2 + \frac{1}{\sigma_{\ln T}} (\mu_{\ln T} - \ln \hat{T})^2 \end{aligned} \tag{3.6}$$

The estimate of $\hat{q}$ and $\hat{T}$ are found by minimising Eq. 3.6.

3.6 Maximum likelihood method

The maximum likelihood estimate is the conditioned density

$$p_{\psi^*|\theta}(\psi^*|\theta) = f_{\psi^*|\theta}(\psi^*|\theta) \tag{3.7}$$

which is commonly called the likelihood function. $f_{\psi^*|\theta}(\psi^*|\theta)$ is the joint probability density function of the observations. The maximum likelihood estimate is a maximisation of Eq. 3.7. In some cases it is convenient to derive the maximum likelihood estimate from a minimisation of the log-likelihood criterion

$$-2 \ln p_{\psi^*|\theta}(\psi^*|\theta) \tag{3.8}$$

If the residuals are assumed to be Gaussian distributed the likelihood function becomes

$$f_{\psi^*|\theta}(\psi^*|\theta) = (2\pi)^{\frac{N}{2}} \mid C_{\psi^*} \mid^{-\frac{1}{2}} e^{\left(\frac{1}{2}(\psi^* - \psi(\theta))^T C_{\psi^*}^{-1}(\psi^* - \psi(\theta))\right)} \tag{3.9}$$

if prior parameter information is included Eq. 3.9 becomes

$$f_{\psi^*|\theta}(\mathbf{z}^*|\theta) = (2\pi)^{\frac{N}{2}} \mid C_{\mathbf{z}^*} \mid^{-\frac{1}{2}} e^{\left(\frac{1}{2}(\mathbf{z}^* - \mathbf{z})^T C_{\mathbf{z}^*}^{-1}(\mathbf{z}^* - \mathbf{z})\right)} \tag{3.10}$$

The parameter set found from minimising 3.10 corresponds to the parameter set found from minimising the least square objective function.

3.7 Single versus multi-objective parameter estimation

Some criticism has been raised against using a single objective function in the estimation of parameters. Yapo et al. (1998) state that *"Practical experiences with model calibration suggests that any single-objective function, no matter how carefully chosen, may not adequately measure the ways in which the model fails to match the important characteristics of the observed data."*

In order to overcome this problem the calibration problem can be formulated as a multi-objective optimisation problem of $F(\theta)$

$$F(\theta) = f_1(\theta), f_2(\theta), ..., f_m(\theta) \tag{3.11}$$

where $f_1(\theta), ..., f_m(\theta)$ are objective functions to be simultaneously minimised with respect to the parameters θ. $f_i(\theta)$ might e.g. be the least square or maximum likelihood parameter estimate with respect to head observations and $f_j(\theta)$

might be the parameter estimate with respect to river discharge observations. A minimisation of the individual objective functions may result in parameter solutions that are not unique and accordingly it is not possible to find the best solution using objective methods.

Figure 3.7 illustrates a problem with two objectives (f_1, f_2) to be minimised with respect to two parameters (θ_1, θ_2). Point A is the solution to minimising f_1 and point B is the solution to minimising f_2

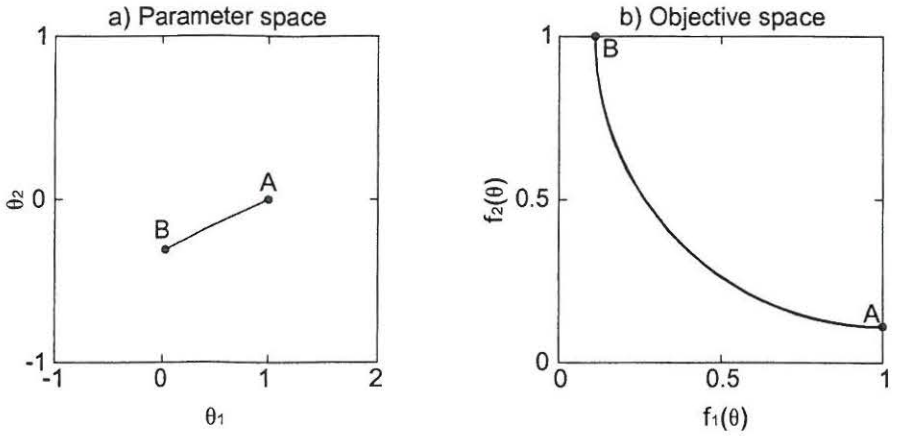


Figure 3.7: Illustration of Pareto optimality. After Yapo et al. (1998)

The solution to the multi-objective problem consists of all parameter combinations on the line from point A to point B. Parameter sets close to point B will result in a small of value of f_1, and as they move towards point A f_1 will decrease and f_2 will increase. Solutions on the line from A to B are called Pareto solutions. (Yapo et al. 1998) All Pareto solutions are acceptable simulators of the system, and the parameter range given by the Pareto solutions reveals the uncertainty due to the choice of objective function. A thorough description of the multi-objective approach and solution methods for minimising multi-objectives can be found in Yapo et al. (1998).

Example 3.3 *Let us consider the flow situation in example 1.1, page 7. Two objective functions are formulated: 1) the least square measure of the head in wells nos. 2 and 3:*

$$J_\psi(\hat{q},\hat{T}) = \frac{1}{\sigma_{\psi_{h_2^*}}}(\psi_2^* - \psi_2(\hat{q},\hat{T}))^2 + \frac{1}{\sigma_{\psi_{h_3^*}}}(\psi_3^* - \psi_3(\hat{q},\hat{T}))^2 \tag{3.12}$$

and 2) the least square measure of the river inflow:

$$J_{Q_r}(\hat{q},\hat{T}) = \frac{1}{\sigma_{\psi_{Q_r^*}}}(Q_r^* - Q_r(\hat{q},\hat{T}))^2 \tag{3.13}$$

The minimisation of J_ψ and J_{q_r} will lead to two estimates of q and T that correspond to points A and B on Figure 3.7. If the two objectives are combined

$$\begin{aligned} J_{\psi,Q_r}(\hat{q},\hat{T}) =& \omega(\psi_2^* - \psi_2(\hat{q},\hat{T}))^2 + (\psi_3^* - \psi_3(\hat{q},\hat{T}))^2 \\ &+ (1-\omega)(Q_r^* - Q_r(\hat{q},\hat{T}))^2 \end{aligned} \tag{3.14}$$

and the weight, ω, is varied from 1 to zero, solutions on the Pareto front can be found.

3.8 Parameter statistics

Parameter statistics can be analysed in case of linear models that have non-biased and Gaussian residuals. The covariance of the estimated parameters can be found from Bard (1974):

$$\mathbf{C}(\hat{\theta}) = \frac{J(\hat{\theta})}{N_{obs} - N_{par}} \left[\mathbf{A^k}(\hat{\theta})\right]^{-1} \tag{3.15}$$

where

$\hat{\theta}$ estimated parameter vector

$J(\hat{\theta})$ least square error

$\mathbf{A^k}(\hat{\theta}) = \left[\mathbf{J_D}\left(\boldsymbol{\theta^k}\right)\right]^T \left[\mathbf{J_D}\left(\boldsymbol{\theta^k}\right)\right], (N_{par} \times N_{par})$

$\mathbf{J_D}$ Jacobian matrix of state variable, ψ, with respect to parameters, θ, $(M \times L)$

$$\mathbf{J_D} = \begin{pmatrix} \frac{\partial\psi_1}{\partial\theta_1} & \frac{\partial\psi_1}{\partial\theta_2} & \cdots & \frac{\partial\psi_1}{\partial\theta_{N_{par}}} \\ \frac{\partial\psi_2}{\partial\theta_1} & \frac{\partial\psi_2}{\partial\theta_2} & \cdots & \frac{\partial\psi_2}{\partial\theta_{N_{par}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\psi_{N_{obs}}}{\partial\theta_1} & \frac{\partial\psi_{N_{obs}}}{\partial\theta_2} & \cdots & \frac{\partial\psi_{N_{obs}}}{\partial\theta_{N_{par}}} \end{pmatrix}$$

N_{obs} number of observations

N_{par} number of parameters

The Jacobian is a local linear approximation of the response surface in the region around the optimal solution, $\hat{\boldsymbol{\theta}}$.

The size of the components of $\mathbf{A^k}$ is proportional to the number of observations and consequently the estimated parameter variation becomes (through Eq.

3.15) inversely proportional to the number of observations. As the number of observations increases the parameter error will diminish.

The linear confidence intervals for the parameter θ_i can be found (Hill 1992, p. 58: Seber and Wild 1989, p. 191-194):

$$\theta_i \pm t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right) s_{\theta_i} \tag{3.16}$$

where $t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right)$ is the Student-t statistic for $N_{obs} - N_{par}$ degrees of freedom and a significance level of α and $s_{\theta_i} = \sqrt{C(\theta_i, \theta_i)}$ is the standard deviation of θ_i.

Example 3.4 *We now want to consider the statistics for the parameter estimate described in example 3.2, page 45. From Eq. 3.15 we have the covariance of the estimated parameters:*

$$\mathbf{C} = \begin{pmatrix} \sigma_{\hat{q}}^2 & \sigma_{\hat{q}}\sigma_{\hat{T}} \\ \sigma_{\hat{q}}\sigma_{\hat{T}} & \sigma_{\hat{T}}^2 \end{pmatrix} = \frac{J(\hat{q}, \hat{T})}{3-2} \left[\mathbf{A^k}(\hat{\boldsymbol{\theta}})\right]^{-1} \tag{3.17}$$

where $\mathbf{A^k}(\hat{\boldsymbol{\theta}}) = \mathbf{J_D}^{\mathbf{T}}\mathbf{J_D}$ *and the Jacobian matrix,* $\mathbf{J_D}$*, is given as*

$$\mathbf{J_D} = \begin{pmatrix} \frac{\partial \psi_2}{\partial \hat{q}} & \frac{\partial \psi_2}{\partial \hat{T}} \\ \frac{\partial \psi_3}{\partial \hat{q}} & \frac{\partial \psi_3}{\partial \hat{T}} \\ \frac{\partial Q_r}{\partial \hat{q}} & \frac{\partial Q_r}{\partial \hat{T}} \end{pmatrix} \tag{3.18}$$

The 95% confidence intervals of $\hat{q}$ *and* $\hat{T}$ *are*

$$\begin{aligned} \hat{q} &\pm t(3-2, 0.975)\,\sigma_{\hat{q}} \\ \hat{T} &\pm t(3-2, 0.975)\,\sigma_{\hat{T}} \end{aligned} \tag{3.19}$$

3.9 State variable statistics

Linear confidence intervals for any system state variables ψ_l are given as (Hill 1992 ,p. 58).

$$\psi_l \pm t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right) s_{\psi_l} \tag{3.20}$$

where s_{ψ_l} is the standard deviation of ψ_l calculated from

$$s_{\psi_l} = \left[\sum_{i=1}^{N_{par}} \sum_{j=1}^{N_{par}} \frac{\partial \psi_l}{\partial \theta_j} C_{\theta_i,\theta_j} \frac{\partial \psi_l}{\partial \theta_i}\right]^{\frac{1}{2}} \tag{3.21}$$

where C_{θ_i,θ_j} are components in the covariance matrix for $\boldsymbol{\theta}$

Approximate linear prediction intervals are calculated as (Hill 1994 ,p. 32)

$$\psi_l \pm t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right)\left(s_{\psi_l}^2 + \frac{s_r^2}{\omega_l}\right)^{1/2} \tag{3.22}$$

where s_r is the calculated standard error of the regression and ω_l is a weight that equals σ_r^2/σ_l^2, where σ_r^2 is the estimated common error variance of the regression, and σ_l^2 is the measurement error variance associated with ψ_l. The error variance of ψ_l is not usually known. Christensen and Cooley (1999) assumed that σ_l^2 was spatially distributed proportionally to the variance of the observed head measurements. Error in predictions becomes equally distributed according to the assumed errors deriving from observation error.

Example 3.5 *We now want to consider the statistics for the estimate of ψ_1, see example 1.1, page 7, and 3.2, page 45. From Eq. 3.21 we have the standard deviation on ψ_1*

$$\begin{aligned} s_{\psi_1} =& \frac{\partial\psi_1}{\partial q} C_{q,T} \frac{\partial\psi_1}{\partial T} + \frac{\partial\psi_1}{\partial q} C_{q,q} \frac{\partial\psi_1}{\partial q} \\ &+ \frac{\partial\psi_1}{\partial T} C_{T,q} \frac{\partial\psi_1}{\partial q} + \frac{\partial\psi_1}{\partial T} C_{T,T} \frac{\partial\psi_1}{\partial T} \end{aligned} \tag{3.23}$$

From Eq. 3.20 the 95 % confidence intervals can be found

$$\psi_1 \pm t\,(3-2, 0.975)\, s_{\psi_1} \tag{3.24}$$

3.10 Solving the regression problem

3.10.1 Gauss-Newton minimisation

The Gauss-Newton algorithm has often been used as a minimisation algorithm in problems concerning the estimation of groundwater parameters. The algorithm starts with an initial parameter vector $\boldsymbol{\theta}^0$ and converges iteratively to a local minimum. The iterative equation and the normal equation used in the Gauss-Newton algorithm can be expressed as (e.g. see Yeh 1986)

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \rho^k \mathbf{d}^k \tag{3.25}$$

with

$$\mathbf{A}^k \mathbf{d}^k = \mathbf{g}^k \tag{3.26}$$

where

$\mathbf{A^k} = \left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]^T \left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]$,$(N_{par} \times N_{par})$;

$\mathbf{g^k} = \left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]^T \left[\psi(\boldsymbol{\theta}^{\mathbf{k}}) - \psi^*\right]$,$(N_{par} \times 1)$;

$\mathbf{J_D}$ Jacobian matrix of state variable, ψ, with respect to parameters, θ, $(M \times L)$

ρ^k scalar step size

$\mathbf{d^k}$ Gauss-Newton direction vector

N_{obs} number of observations

N_{par} number of parameters

A description of the modified Gauss-Newton Optimisation Method, including weighted residuals, can be found in Hill 1992, p. 76-82.

3.10.2 Other methods

An alternative to the Gauss-Newton minimisation is offered by global random search methods, where the parameter space is sampled randomly and global parameter set/sets are estimated on the basis of one or more objective functions.

Other algorithms combine local search methods with random search methods. The Multi-Objective Complex Evolution (MOCOM-UA) method (Yapo et al. 1998) and the shuffled complex evolution (SCE-UA) method (Duan et al. 1992) are examples of such methods. Madsen and Kristensen (2002) applied UCODE (inverse programme for Gauss-Newton minimisation) and the SCE-UA method on a MIKE SHE application and found that "The UCODE solutions were trapped in local optima far from the Pareto front. Even when the initial parameter set was close to the Pareto front, UCODE failed to converge into a Pareto optimal solution."

CHAPTER 4

The Generalized Likelihood Uncertainty Estimation methodology

Calibration and uncertainty estimation based upon a statistical framework is aimed at finding an optimal set of models, parameters and variables capable of simulating a given system.

There are many possible sources of mismatch between observed and simulated state variables (see section 3.2). Some of the sources of uncertainty originate from physical randomness, and others from uncertain knowledge put into the system. The uncertainties originating from physical randomness may be treated within a statistical framework, whereas alternative methods may be needed to account for uncertainties originating from the interpretation of incomplete and perhaps ambiguous data sets.

The GLUE methodology (Beven and Binley 1992) rejects the idea of one single optimal solution and adopts the concept of equifinality of models, parameters and variables (Beven and Binley 1992; Beven 1993). Equifinality originates from the imperfect knowledge of the system under consideration, and many sets of models, parameters and variables may therefore be considered equal or almost equal simulators of the system. Using the GLUE analysis, the prior set of models, parameters and variables is divided into a set of non-acceptable solutions and a set of acceptable solutions. The GLUE methodology deals with the variable degree of membership of the sets. The degree of membership is determined by assessing the extent to which solutions fit the model, which in turn is determined

by subjective likelihood functions. By abandoning the statistical framework we also abandon the traditional definition of uncertainty and in general will have to accept that to some extent uncertainty is a matter of subjective and individual interpretation by the hydrologist. There are strong parallels between uncertainty in a Fuzzy set ruled system and uncertainty in the GLUE methodology. Fuzzy logic is an alternative or supplement to the classical probabilistic framework in situations where very little information is available, and such information as there is tends to be ambiguous and vague. Considering the sources of mismatch between observed and simulated state variables (see section 3.2), it can be argued that the mismatch is to a great extent due to vague and ambiguous interpretations.

The GLUE methodology consists of the 3 steps described below (Fig. 4.1).

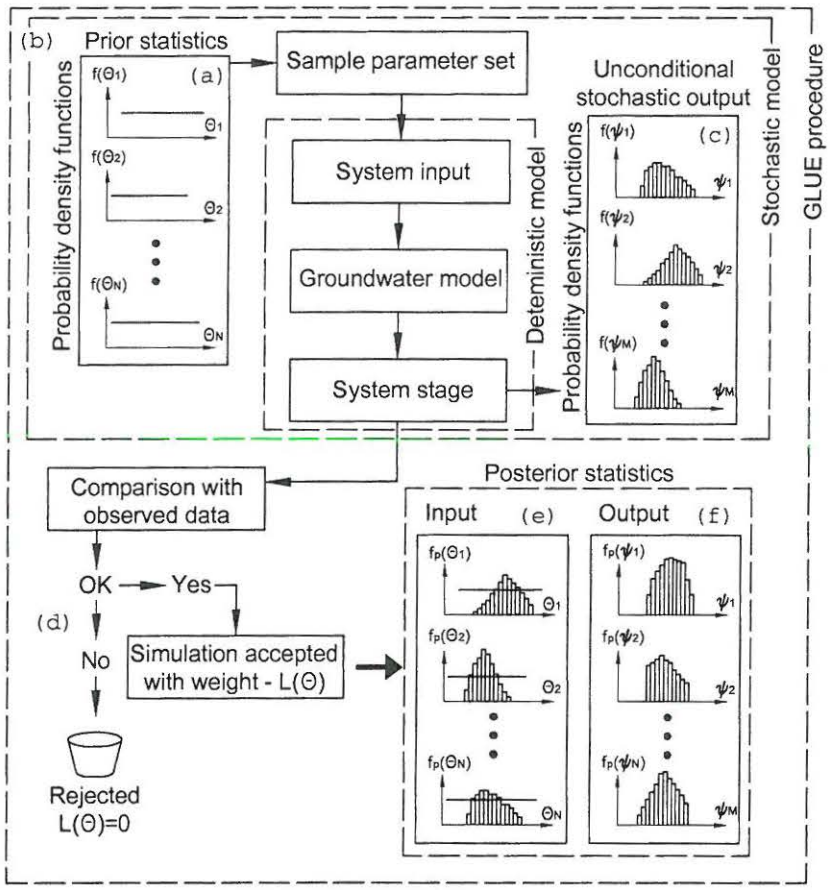


Figure 4.1: The GLUE procedure. (a) prior statistics, (b) stochastic modelling, (c) unconditional statistics of system state variables, (d) evaluation procedure (e) posterior parameter likelihood functions and (f) likelihood functions for system state variables

Step 1 is to determine the statistics for the models, parameters and variables

that, prior to the investigation, are considered likely to be decisive for the simulation of the system **(a)**. Typically quite wide discrete or continuous uniform distribution is chosen - reflecting the fact that there is little prior knowledge of the uncertainties arising from models, parameters and variables. In principle all available knowledge can be put into the prior distributions.

Step 2 is a stochastic simulation **(b)** based on the models, parameters and variables defined in step 1. The Monte Carlo or Latin Hypercube method (Appendix A) may be used to do a random sample of the parameter sets. Step 2 gives us an unconditional estimate of the statistics of any system state variable **(c)**.

In step 3 an evaluation procedure **(d)** is carried out for every single simulation performed in step 2. Simulations and thus parameter sets are rated according to the degree to which they fit observed data. If the simulated state variables are "close" to the observed values the simulation is accepted as having a given likelihood $L(\boldsymbol{\theta}|\boldsymbol{\psi})$, whereas if the considered simulated state variables are unrealistic the simulation is rejected as having zero likelihood.

In this way a likelihood value is assigned to all accepted parameter sets (zero for rejected sets and positive for accepted sets). The direct result of this is a discrete joint likelihood function (DJPDF) for all the models, parameters and variables involved. The DJPDF can only be illustrated in two, maximum three, dimensions, and likelihood scatter plots are often used to illustrate the estimated parameters, see e.g. Fig. 5.7. In Fig. 4.1 the models, parameters and variables $\theta_1, ..., \theta_i, ..., \theta_N$ are considered independent, the likelihood is projected onto the parameter axis, and discrete density functions **(e)** are presented, see section 4.3. Discrete likelihood functions for all types of system state variables can likewise be constructed **(f)**.

4.1 Likelihood measures

Likelihood is a measure of how well a given combination of models, parameters and variables fits, based on the available set of observations. The likelihood measure thus describes the degree to which the various acceptable solutions are members of the set, i.e. their degree of membership.

The calculation of the likelihood of a given set of models, parameters and variables is the key feature of the GLUE methodology, and in this respect GLUE differs from the classical methods of calibration and uncertainty estimation. As will be seen in what follows a wide range of likelihood measures are suggested - all with different qualities. There are no definitive rules for choosing a certain likelihood measure. Some personal preferences are however mentioned in section 4.2.

The likelihood measure consists, in this thesis, of three elements: 1) a rejection level that indicates whether the acceptance criteria are fulfilled or not, 2) a point likelihood measure that sums up the degree of model fit in the individual observation points and 3) a global likelihood measure that is an aggregation of all the point likelihood measures.

Often the rejection level is implicitly given in the point likelihood function, and occasionally the rejection level, the point likelihood measure and the global likelihood measure are all gathered in one function.

The likelihood functions presented below in Fig. 4.2 are based on a combination of the likelihood functions derived from the classical statistical framework and from GLUE, and the Fuzzy logic literature.

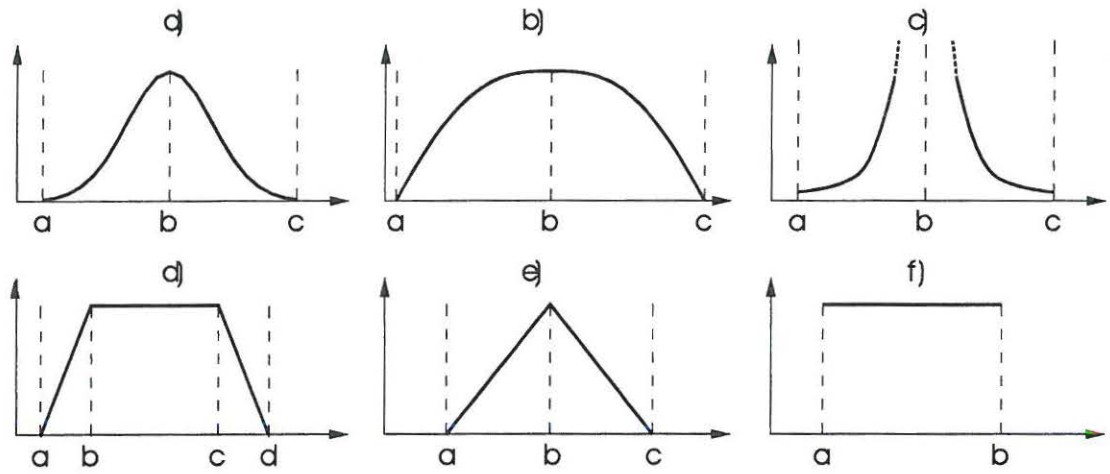


Figure 4.2: a) Gaussian likelihood function , b) model efficiency likelihood function, c) inverse error variance likelihood function, d) trapezoidal likelihood function, e) triangular likelihood function and f) uniform likelihood function

4.1.1 Traditional statistical likelihood measures

Gaussian likelihood function

The Gaussian likelihood function, Fig. 4.2a, is often used in a classical statistical framework. The residuals are assumed to be Gaussian and the likelihood equals the probability that the simulated value, $\psi_i(\boldsymbol{\theta})$, equals the observed value, ψ_i^*:

$$L\left(\boldsymbol{\theta}|\psi_i^*\right) = \frac{1}{\sqrt{2\pi}\sigma_{\psi_i^*}} e^{-\left(\frac{\left(\psi_i^* - \psi_i(\boldsymbol{\theta})\right)^2}{2\sigma^2_{\psi_i^*}}\right)} \tag{4.1}$$

or for N_{obs} observations

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = (2\pi)^{\frac{-N_{obs}}{2}} \mid C_{\boldsymbol{\psi}^*} \mid^{-\frac{1}{2}} e^{\left(\frac{1}{2}(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta}))^T C^{-1}_{\boldsymbol{\psi}^*}(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta}))\right)} \tag{4.2}$$

where, as in a statistical framework $\sigma_{\psi_i^*}$ and $C_{\boldsymbol{\psi}^*}$ symbolise the unknown standard deviation and covariance of observed state variables - often approximated by the expected standard deviation and covariance of observed state variables. Eq. 4.2 corresponds to the product inference function (section 4.1.4) of Eq. 4.1, given independent observations.

The N-dimensional Gaussian likelihood function (4.2) is a function that depends on the number of observations. As the number of observations increases, so does the likelihood of the best simulations, until finally ($N_{obs} \rightarrow \infty$) all likelihood is ascribed to the single best simulation. The likelihood function yields parameter and uncertainty estimates that are similar to those achieved within a statistical framework when this is applied to well-posed linear models with Gaussian errors and the estimate implicit assumes that the model is error free.

The Gaussian likelihood function is defined from $-\infty$ to ∞ and thus no rejection level is implicitly given. In order to reduce the number of accepted simulations, it will often be appropriate to introduce a rejection level (a and c on Fig. 4.2a), e.g. at three times the standard deviation.

4.1.2 Traditional GLUE likelihood measures

Model efficiency function

The model efficiency function, Fig. 4.2b, is given as (Beven and Binley 1992)

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \left(1 - \sigma_\varepsilon^2/\sigma_0^2\right); \quad \sigma_\varepsilon^2 \geq \sigma_0^2 \Rightarrow L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = 0 \tag{4.3}$$

where

$$\sigma_\varepsilon^2 = \frac{1}{N_{obs}} \left(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})\right)^T \mathbf{V} \left(\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})\right) \tag{4.4}$$

is the weighted variance of the residuals and σ_0^2 is the weighted variance of the observations. Here $\mathbf{V}$ is a weight matrix.

The likelihood equals one if all residuals are zero, and zero if the weighted variance of the residuals is larger then the weighted variance of the observations.

Inverse error variance function

Beven and Binley (1992) have suggested a function based on the inverse error variance with shaping factor N, Fig. 4.2c:

$$L\left(\boldsymbol{\theta}|\psi^*\right) = \left(\sigma_\varepsilon^2\right)^{-N} \tag{4.5}$$

This function concentrates the weights of the best simulations as N increases. For $N \to \infty$ all weight will be on the single best simulation and for small values of N all simulations will tend to have equal weight.

4.1.3 Fuzzy likelihood measures

A point observation of the i^{th} system state variable, ψ_i^*, and a computed value of the same system state variable, $\psi_i(\boldsymbol{\theta})$ are considered. In the set of all possible values of ψ_i, a subset, Ψ_i, is defined where the transition between membership and non-membership is gradual. The likelihood - or, in Fuzzy terms, the degree of membership - is maximum for simulated state variables that belong completely to Ψ_i; elsewhere it is between 0 and the maximum value. In Fuzzy logic Ψ_i is called a fuzzy set and the likelihood (degree of membership) is described by the likelihood function (membership function), $Ł_{\Psi_i}$. The likelihood function can in principle be an arbitrary, non-symmetric and biased function. The trapezoidal, triangular and uniform likelihood functions are typical Fuzzy logic membership functions where the likelihood or degree of membership is evaluated through relatively simple functions.

First the point likelihood measures are described, and then the point likelihood measures are combined through the so-called inference functions.

Trapezoidal likelihood function

The trapezoidal likelihood function, Fig. 4.2d, is given as

$$L\left(\boldsymbol{\theta}|\psi_i^*\right) = \frac{\psi_i(\boldsymbol{\theta}) - a}{b-a} I_{a,b}(\psi_i(\boldsymbol{\theta})) + I_{b,c}(\psi_i(\boldsymbol{\theta})) + \frac{d - \psi_i(\boldsymbol{\theta})}{d-c} I_{c,d}(\psi_i(\boldsymbol{\theta})) \tag{4.6}$$

where

$$I_{a,b} = \begin{cases} 1 & \text{if } a \leq \psi_i(\boldsymbol{\theta}) \leq b \\ 0 & \text{otherwise} \end{cases}$$
$$I_{b,c} = \begin{cases} 1 & \text{if } b \leq \psi_i(\boldsymbol{\theta}) \leq c \\ 0 & \text{otherwise} \end{cases}$$
$$I_{c,d} = \begin{cases} 1 & \text{if } c \leq \psi_i(\boldsymbol{\theta}) \leq d \\ 0 & \text{otherwise} \end{cases}$$

Triangular likelihood function

The triangular likelihood function, Fig. 4.2e, is given as

$$L(\boldsymbol{\theta}|\psi_i^*) = \frac{\psi_i(\boldsymbol{\theta}) - a}{b - a} I_{a,b}(\psi_i(\boldsymbol{\theta})) + \frac{c - \psi_i(\boldsymbol{\theta})}{c - b} I_{b,c}(\psi_i(\boldsymbol{\theta})) \tag{4.7}$$

where

$$I_{a,b} = \begin{cases} 1 & \text{if } a \leq \psi_i(\boldsymbol{\theta}) \leq b \\ 0 & \text{otherwise} \end{cases}$$
$$I_{b,c} = \begin{cases} 1 & \text{if } b \leq \psi_i(\boldsymbol{\theta}) \leq c \\ 0 & \text{otherwise} \end{cases}$$

Uniform likelihood function

The uniform likelihood function, Fig. 4.2f, is a special case of the trapezoidal likelihood function where $a = b$ and $c = d$.

$$L(\boldsymbol{\theta}|\psi_i^*) = \begin{cases} 1 & \text{if } a < \psi_i^* - \psi_i(\boldsymbol{\theta}) < b \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

4.1.4 Inference functions

The overall combination of the individual point likelihood (degree of membership) for the observation points is assembled through the so-called degree of fulfilment (DOF) (Dubois and Prade 1980), which, in this context, is the overall likelihood value for the simulation - a global likelihood measure, $L(\boldsymbol{\theta}|\boldsymbol{\psi}^*)$. A classification of aggregation operators used in Fuzzy rules systems is given in Zimmermann (1991), p. 40-41, and some relevant operators are given below (Dubois and Prade 1980; Zimmermann 1991):

Product inference

$$L(\boldsymbol{\theta}|\boldsymbol{\psi}^*) = \prod_{i=1}^{N_{obs}} L(\boldsymbol{\theta}|\psi_i^*) \tag{4.9}$$

The product inference is very restrictive - if one observation is outside the Fuzzy set, Ψ, (i.e. rejected) the global likelihood will be zero. As N_{obs} increases, the global likelihood response surface becomes steeper and steeper and as $N_{obs} \rightarrow \infty$ all except the single best simulation will have negligible likelihood.

Min. inference

$$L(\boldsymbol{\theta}|\boldsymbol{\psi}^*) = \min_{i=1,\ldots,N_{obs}} L(\boldsymbol{\theta}|\psi_i^*) \tag{4.10}$$

The min. inference is as restrictive as the product inference function but the global likelihood response surface is more flat.

Max. inference

$$L(\boldsymbol{\theta}|\boldsymbol{\psi}^*) = \max_{i=1,\ldots,N_{obs}} L(\boldsymbol{\theta}|\psi_i^*) \tag{4.11}$$

The max. inference is the least restrictive inference function. The likelihood is evaluated from the observation point with the best agreement. If just one observation is inside the Fuzzy set (i.e. accepted), then the simulation is accepted.

Weighted arithmetic mean inference

$$L(\boldsymbol{\theta}|\boldsymbol{\psi}^*) = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \omega_i L(\boldsymbol{\theta}|\psi_i^*) \tag{4.12}$$

where ω_i is the weight on the i^{th} observation.

As in the case of max. inference, the inclusion of just one observation within the accepted set will result in acceptance of the simulation. The response surface for the arithmetic mean inference is very flat.

Geometric mean inference

$$L(\boldsymbol{\theta}|\psi^*) = \sqrt[N_{obs}]{\prod_{i=1}^{N_{obs}} L(\boldsymbol{\theta}|\psi_i^*)} \tag{4.13}$$

The geometric mean inference is as restrictive as the product and min. inference, but the likelihood response surface is less steep. The function is independent of the number of observations.

The way that a likelihood calculation might be performed when different types of observation data are available is illustrated in example 4.1 below.

Example 4.1 *A stationary groundwater model is constructed for a river catchment. The model is calibrated to a summer situation. The following observations are available:*

- *Head observations in 16 wells. From initial studies the standard error on the observed heads is estimated to be 1.5 m. Trapezoidal likelihood functions are applied. Fig. 4.3(a)*
- *Median value of annual minimum discharge observations at one station in "Large Creek". The estimation error is assumed to be Gaussian with a standard error of 10 % of measured discharge. The rejection level is three times standard error. Fig. 4.3(b)*
- *A local farmer has stated that "Little Creek dries out every summer". We do not rely totally on this statement and formulate a likelihood function that gradually decreases from 0 l/s to 2.0 l/s. Fig. 4.3(c)*
- *Information from the local waterworks indicates that so far abstraction well no. 12 has never dried out. Low hydraulic conductivities may result in the closing of abstraction wells in the numerical model. Seen in the light of the information given above, every simulation where the abstraction is closed must be unrealistic, and consequently the likelihood is set at zero. Fig. 4.3(d)*

In all, 19 observations are available and they are combined into an global simulation likelihood measure by an inference rule, e.g. weighted arithmetic mean, Eq. 4.12, or geometric mean inference, Eq. 4.13. Alternatively two or more rules can be combined, e.g. Eq. 4.14.

$$\begin{aligned} L(\boldsymbol{\theta}|\psi^*) = \; & \omega_{head} \tfrac{1}{16} \textstyle\sum_{i=1}^{16} L_{h_i^*}(h_i(\boldsymbol{\theta})) \cdot \omega_{q_1} L_{q_1^*}(q_1(\boldsymbol{\theta})) \\ & \cdot \omega_{q_2} L_{q_2^*}(q_2(\boldsymbol{\theta})) \cdot \omega_{Abs} L_{Abs_2^*}(Abs(\boldsymbol{\theta})) \end{aligned} \tag{4.14}$$

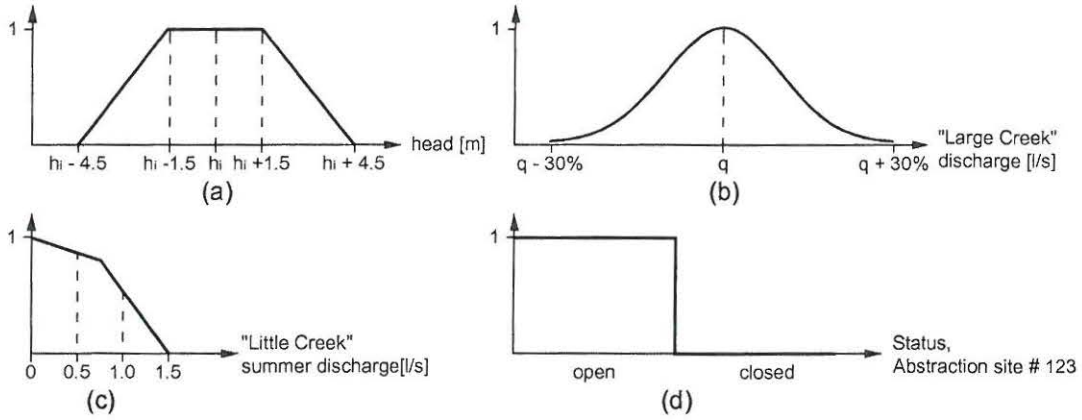


Figure 4.3: Examples of different likelihood functions

where ω_{head}, ω_{q_1}, ω_{q_2} and ω_{Abs} are the weight on observed head data, observed discharge in "Little Creek", observed discharge in "Large Creek" and waterworks observation respectively. $L_{h_i^*}(h_i(\boldsymbol{\theta}))$, $L_{q_1^*}(q_1(\boldsymbol{\theta}))$, $L_{q_2^*}(q_2(\boldsymbol{\theta}))$ and $L_{Abs_2^*}(Abs_2(\boldsymbol{\theta}))$ are likelihood functions for head data, discharge in "Little Creek", discharge in "Large Creek" and the abstraction respectively.

4.2 Designing the likelihood measure

The GLUE methodology is aimed at finding possible sets of models, parameters and variables which produce a model output that is in agreement with observations. The likelihood measure reflects the degree to which we accept the simulated output to deviate from observations due to the numerous error sources.

The first step in the construction of the likelihood function is to analyse possible sources of mismatch between observed and simulated state variables. Section 3.2 is a description of the different types of observation data and a description of the different sources of mismatch between observed and simulated values. Section 3.2 may be used as a guideline in estimating the expected standard errors of observation. In reviewing the possible errors, the hydrologist is forced to consider what is included in the model and what is not. E.g. if the purpose of the model is to model small-scale point pollution, small-scale heterogeneity is very important and consequently has to be modelled in such a way that the error contribution from ignored small-scale heterogeneities will be very small.

In the opinion of the author the estimated expected error should be closely related to the likelihood measure. The rejection level may be three times the expected standard error, reflecting a very low probability of larger errors: see

Chapters 5 and 6

The second step in the calculation of the likelihood measure is the combination of the individual point likelihood measures into a global likelihood measure.

The aim of the point likelihood measures is to account for all expected uncertainty, and in the author's opinion therefore the simulation can be accepted only if all point observations are accepted - no simulated state variables can be tolerated outside the rejection level. If this results in an over-restrictive likelihood measure, the point likelihood measures, and thus the expected errors, should be reconsidered, and if there is no objective reason for increasing the amount of expected error the model should be reconsidered.

The min. inference, the product inference and the geometric mean inference function fulfil the requirement listed above (all point likelihood measures have to be positive in order to accept the simulation).

The geometric mean inference function is attractive because the likelihood measure is independent of the number of observations. This means that the uncertainty estimate does not improve if the number of observations is doubled. This behaviour contrasts with the classical regression framework, where it is assumed that the estimation error is reduced as the number of observations increases. Actually, the maximum likelihood estimate for N independent parameters is the product inference of the independent maximum likelihood estimate.

The reason why the geometric mean inference function is found attractive lies within the error sources. From section 3.2 it can be seen that the main error contributions (scale errors) do not disappear as the number of observations increases, and neither should the uncertainty of the model outcome.

Following the GLUE analysis a validation of all observation points should be performed. From the accepted simulations the probability density functions for the simulated values in the observations points can be found, and the majority of the observations should be within the 95% prediction interval. A poor validation indicates that the likelihood measure is too restrictive and that not all sources of uncertainty are accounted for. See sections 5.6.8 and 6.4.

4.3 Bayesian updating of prior parameter distributions

Following the GLUE analysis the likelihoods are known in a number of discrete points in the space of models, parameters and variables. The posterior likelihood functions for the models, parameters and variables involved can be found from

Bayes' theorem

$$L_p(\boldsymbol{\theta}|\boldsymbol{\psi}^*) = \frac{L(\boldsymbol{\theta}|\boldsymbol{\psi}^*)\, L(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\boldsymbol{\psi}^*)\, L(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \tag{4.15}$$

where $L_p(\boldsymbol{\theta}|\boldsymbol{\psi}^*)$ is the posterior likelihood distribution for models, parameters and variables and $L(\boldsymbol{\theta})$) is the prior likelihood/probability distribution for models, parameters and variables.

Let us for example assume the we have N_{acc} acceptable parameter sets with likelihood $L(\boldsymbol{\theta}_1|\boldsymbol{\psi}^*), ..., L(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*) ..., L(\boldsymbol{\theta}_{N_{acc}|\boldsymbol{\psi}^*})$ and from the joint prior likelihood /probability distribution we have corresponding prior likelihood at the same points in parameter space $L(\boldsymbol{\theta}_1), ..., L(\boldsymbol{\theta}_i) ..., L(\boldsymbol{\theta}_{N_{acc}})$ The posterior likelihood of the points considered in the space of models, parameters and variables is

$$L_p(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*) = \frac{L(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*)\, L(\boldsymbol{\theta}_i)}{\sum_{i=1}^{N_{acc}} L(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*)\, L(\boldsymbol{\theta}_i)} \tag{4.16}$$

It can be shown that in the case of uniform prior distributions the posterior likelihood equals the GLUE computed likelihood, $L_p(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*) = L(\boldsymbol{\theta}_i|\boldsymbol{\psi}^*)$.

4.4 An example

In example 3.1, p. 42, it was argued that both head and river inflow observations were necessary in order to make the calibration of q and T unique. The GLUE methodology does not set restrictions on the basis of uniqueness - non-uniqueness will simply result in a larger range of possible parameter values.

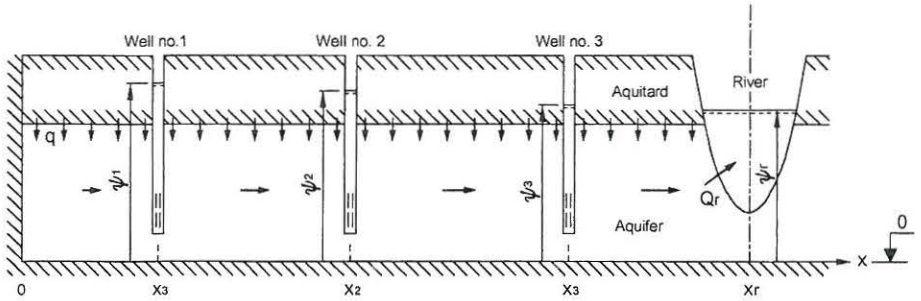
The GLUE methodology is applied to example 1.1, p. 42, with the parameters presented in Fig. 4.4.

A Monte Carlo simulation is performed with 20,000 random realisations of q and T. Each realisation results in an estimate of h_2, h_3 and Q_r. Q_r is found as the total amount of water infiltrated into the aquifer, $Q_r = q \cdot 1000m \cdot 1m$

We now want to use the "observations" of h_2^*, h_3^* and Q_r^* in order to calculate the likelihood of each of the 20,000 simulations. h_2^*, h_3^* and Q_r^* are found from Eq. 1.3 with the parameters:

q $= 400$ mm year^{-1}

T $= 5.\ 10^{-4}$ m^2 s^{-1}



$x_1 = 250$ m $\quad x_2 = 500$ m $\quad x_3 = 750$ m $\quad x_r = 1000$ m $\quad \psi_r = 20$ m
q = U[200,600] (mm year^{-1}) $\quad \log_{10} T$ = U[-3,-4] ($\log_{10}$ (m^2 s^{-1}))

Figure 4.4: Groundwater flow problem and parameters. $U[\cdot]$ denotes uniform distribution

and error of -0.1 m, 0.7 m and 7.34 10^{-7} m^3 s^{-1} are added to h_2^*, h_3^* and Q_r^*, respectively in order to represent observation errors and model errors. This yields

$h_2^* = 29.41$ m

$h_3^* = 26.25$ m

$Q_r^* = 1.332^{-5}$ m^3 s^{-1}

Prior to the simulation the expected standard error in the observations is estimated at 0.3 m on the head observations and 10 % of the observed river inflow.

The trapezoidal point likelihood function is used in the evaluation of h_2^*, h_3^* and Q_r^*, see Fig. 4.5.

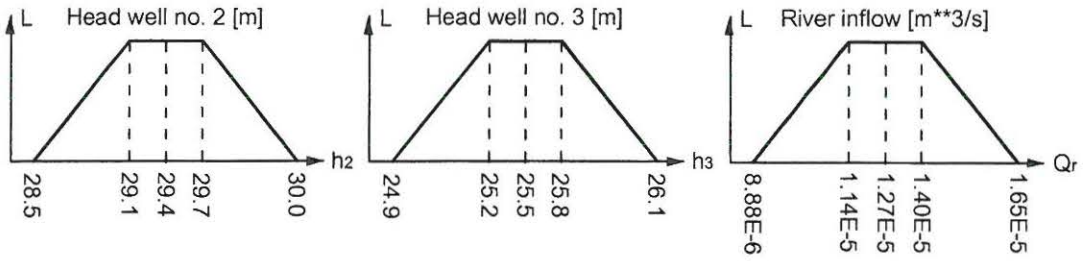


Figure 4.5: Likelihood functions for h_2, h_3 and Q_r

Three point likelihood values, $L_{h_{2,i}}$, $L_{h_{3,i}}$, $L_{Q_{r,i}}$, are calculated on the basis of $h_{2,i}$, $h_{3,i}$ and $Q_{r,i}$ and the global likelihood for the i^{th} simulation is calculated using the geometric mean inference function.

Two scenarios are considered:

a) Only head observations are used in the calculation of the global likelihood:

$$L_i\left(\psi_2^*, \psi_3^* | q_i, T_i\right) = \sqrt{L_{h_{2,i}} L_{h_{3,i}}} \tag{4.17}$$

In this scenario 1,800 of the 20,000 simulations are accepted.

b) Head and river inflow observations are used in the calculation of the global likelihood:

$$L_i\left(\psi_2^*, \psi_3^*, Q_r^* | q_i, T_i\right) = \sqrt[3]{L_{h_{2,i}} L_{h_{3,i}} L_{Q_{r,i}}} \tag{4.18}$$

Here 1,400 of the 20,000 simulations are accepted.

In Figure 4.6 the parameter respond surface for scenarios **a** and **b** is presented.

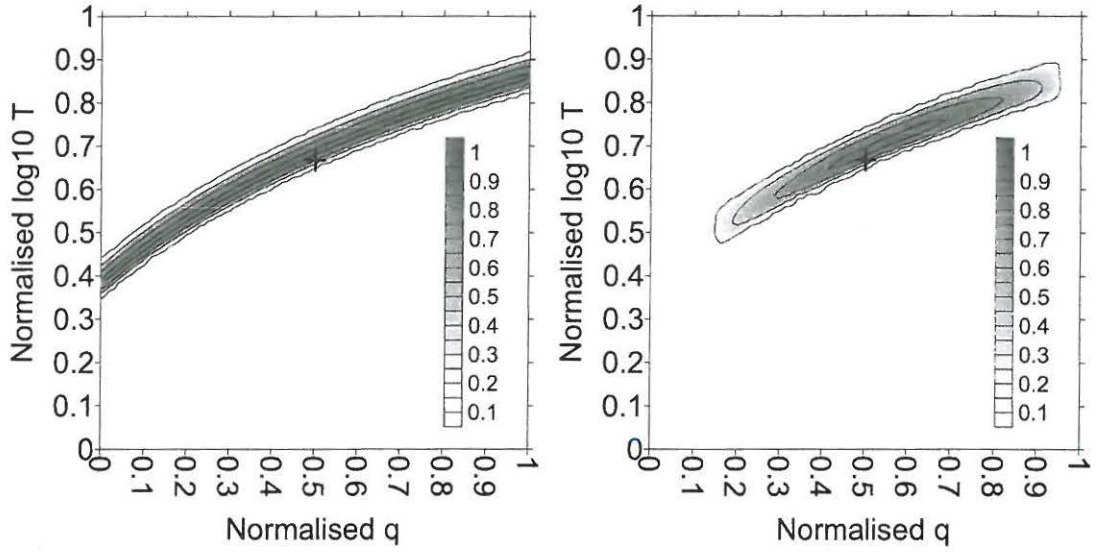


Figure 4.6: Likelihood surfaces in normalised parameter space. ***a)*** *h_2 and h_3 have been used in the calculation of likelihood surface.* ***b)*** *h_2, h_3 and Q_r have been used in the calculation of the likelihood surface. The cross indicates the parameter set used in the calculation of the "observed" values.*

Non-uniqueness is recognised in scenario **a** where only head data are used in the GLUE analysis. If we look at the response surface/curve at a given value of q it is seen that the band of possible T values is quite narrow, but when we look at the total variation of T for all values of q the band is much wider.

In scenario **b** both head and river data are included and the band of possible q values is narrower than in case **a**.

For both scenarios the "true" parameter solution is among the accepted solutions, but not in the region with maximum likelihood. This is due to errors introduced on the observations. If we remove the errors the true parameter solution will fall on the line with maximum likelihood.

The posterior parameter likelihood distributions are identical to the parameter likelihood distribution, because the prior likelihood distribution of q and T is uniform.

If we look at the likelihood distribution curves for ψ_1 for scenarios **a** and **b** we see that they are very similar, Fig. 4.7

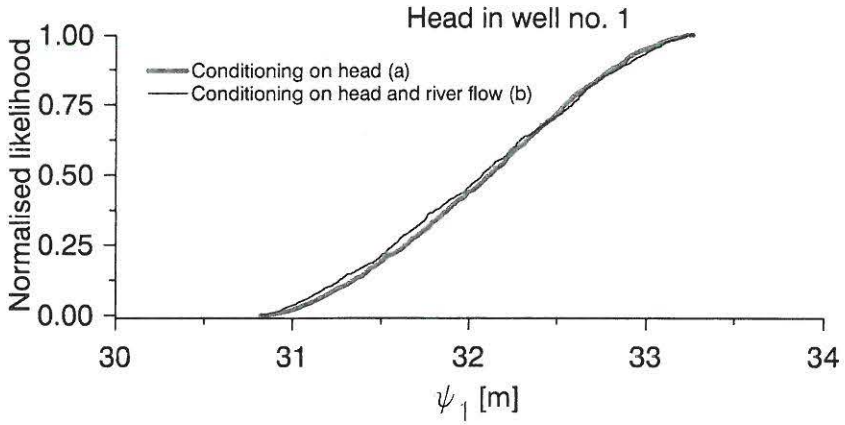


Figure 4.7: Likelihood distribution curves ψ_1.

This indicates that the predictive uncertainty of ψ_1 is mainly influenced by the head observations. If the head rejection criteria are tightened (less expected error in ψ_2 and ψ_3 or a different likelihood function) then there will be less predictive uncertainty in ψ_1. However, this does not mean that predictive capability is invariant to Q_r in general.

Fig. 4.8 presents the likelihood distribution curves of the average Darcy velocity in the aquifer.

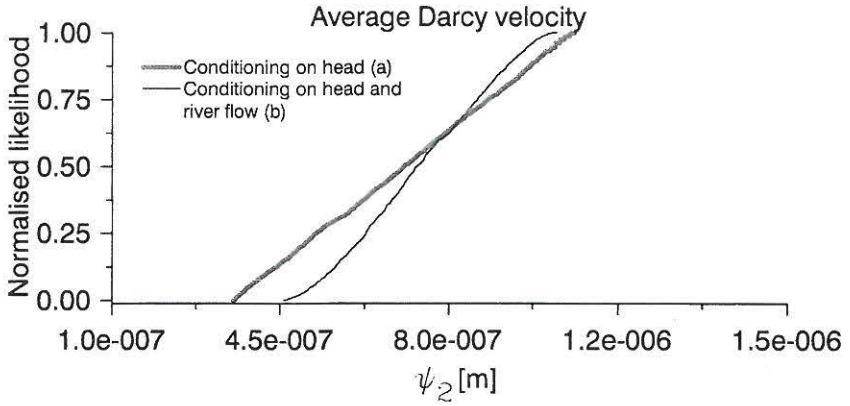


Figure 4.8: Likelihood distribution curves ψ_2.

Scenario **a** results in a significantly larger uncertainty in flow velocities in the aquifer than scenario **b**.

4.5 Generation of random parameter sets

The aim of the GLUE methodology is to find regions in the parameter space resulting in acceptable simulations. Random search methods such as the Monte Carlo method and the Latin Hypercube method have been used in the search in most GLUE applications. These methods are in many ways ineffective because regions of interest often only constitute a small fraction ($< 1\%$) of the prior-defined space of models, parameters and variables. The response surface however is often very complex, with multiple local maxima, valleys and plateaus in a high dimensional parameter space. This makes more intelligent search methods complicated and in some cases inefficient.

In the Gjern setup presented in Chapter 6 an attempt was made to reject certain parameter sets prior to the simulation simply by examining the likelihood in the surrounding region of the parameter set in the parameter space. A similar procedure was used in the original Beven and Binley (1992) study.

The procedure was 1) to generate a parameter set, 2) to interpolate the likelihood value from the surrounding, already simulated, parameter sets, 3) to add a distance-related error to the interpolated value (the closer the point is to the previously sampled parameter sets, the more certain is the interpolation and vice versa) and 4) to simulate the parameter set if the likelihood value was above a certain level. To start with almost all parameter sets were simulated because of the sparse representation, but once a few millions parameter sets had been simulated, up to 60 % of new parameter sets were rejected in advance. There were however no computational benefits from this, due to the costs of interpolating among millions of parameters sets in an 11-dimensional space.

4.6 Concluding remarks

This chapter describes the GLUE methodology that has become a central part of this ph.d. thesis. The use of likelihood functions to evaluate model fit is the key feature of the GLUE methodology. As a supplement to the traditional likelihood measures a number of subjective likelihood measures are introduced and it is thus accepted that the GLUE methodology does not yield uncertainty measures comparable to those produced within the classical statistical framework, but rather offers a statistical measure relating to the subjective impressions of the hydrologist involved. In section 4.2 a few guidelines regarding the design of the likelihood measure have been suggested.

In the following two chapters the GLUE methodology is applied to a synthetic groundwater model and to a region aquifer system.

The main questions to be answered in these two chapters are:

(i) Is it possible from a computational point of view to conduct a GLUE analysis on a typically stationary groundwater model application?

(ii) Is it possible to use the guidelines presented in section 4.2 to design likelihood measures that yield reasonable results?

CHAPTER 5

Case A: Capture zone modelling in a synthetic setup

The estimation of capture zones is a very important issue in the field of groundwater modelling. The delineation of capture zones constitutes the basis for land use regulation aimed at reducing e.g. nitrate and pesticide land surface load, and for prioritising the remediation of polluted sites.

Some work has been done on stochastic capture zone modelling in 2-dimensional groundwater models; see Leeuwen (2000) for a review. Leeuwen's study focused on estimations conditioned on transmissivity data. Feyen et al. (2001) used the Generalized Likelihood Uncertainty Estimation methodology, in which capture zone estimation is conditioned on head observations, and Feyen et al. (2002) include head and transmissivity data in the their work on estimation capture zones.

Interactions between different aquifers and between surface and sub-surface waters may have considerable influence on the delineation of the capture zone. It is therefore essential to quantify the uncertainties in more complex hydrological systems.

This chapter presents a study of a synthetic 3D groundwater system, including surface flow. The aim of the study was to construct a setup containing some of the components and constraints of a real study area. The GLUE methodology was used to condition the capture zone estimates on head and river discharge observations. Hydrological conductivities, net precipitation and the geological model were considered unknowns.

Definition

A **capture zone** of an abstraction well or an abstraction site is the area from which water is captured within time t. Except where otherwise stated, the term capture zone is the time-related capture zone at infinite time.

5.1 Conceptual models

The study area was a 2000 m x 3000 m rectangular river catchment containing a 2200 m long river. The sub-surface flow region consisted of an upper aquifer, an aquitard and a lower aquifer from which groundwater was abstracted. Two equally likely geological models were defined. In the first geological model (A) all three layers were distributed throughout the entire catchment, while in the second geological model (B) a sandy window existed in a 300 m x 2000 m zone in the aquitard, Fig. 5.1. The hydraulic conductivities in the three layers were considered homogeneous.

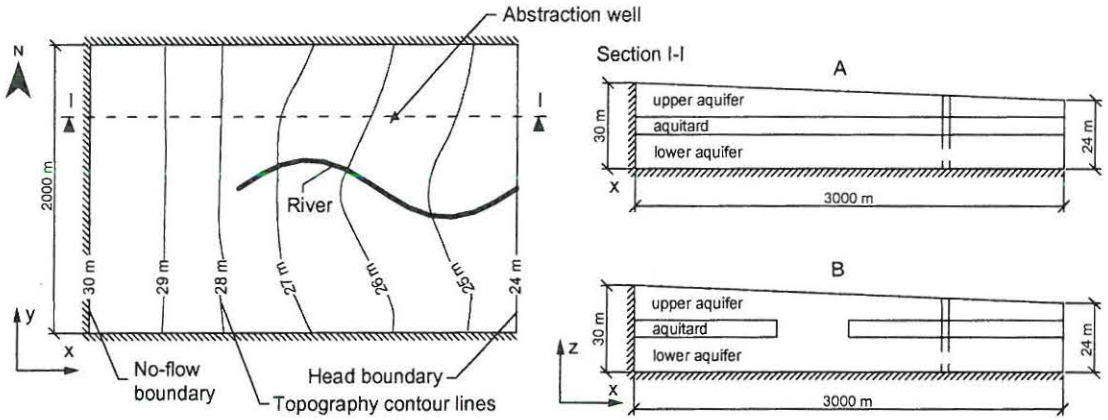


Figure 5.1: Conceptual model.

The study area was bounded by no-flow boundary conditions on the south, west and north boundaries. On the eastern boundary a constant head boundary condition was present with a level of 23.5 m for both the groundwater component and the river component. Groundwater was abstracted at a rate of 50 mm year^{-1} (300,000 m^3 year^{-1}) from the lower aquifer, Fig. 5.1. The net precipitation was added uniformly to the entire catchment.

The river cross section was v-shaped and the Manning numbers were 10 m$^{1/3}$ s^{-1} for the river and 40 m$^{1/3}$ s^{-1} for surface flow.

5.2 Reference model

For calibration purposes a reference model was constructed in order to generate the "observed" data and a reference capture zone. The geometry of this model was identical to the conceptual model presented in Fig. 5.1. The second geological model (B - sandy window) was used and the horizontal hydraulic conductivities in the upper and lower aquifer, and the vertical hydraulic conductivity in the aquitard, were modelled as a random field within the geological layers. The purpose of modelling conductivities as random fields was to introduce model errors in the conceptual description of the model area. In this way the reference model contained geological heterogeneities that were not described in the conceptual model. This corresponds to the situation in most groundwater model applications, where lack of data forces the hydrologist to simplify the conceptual model despite the probable presence of heterogeneities. In generating the random fields, it was assumed that the conductivities were lognormal distributed with an exponential decaying correlation structure.

$$\rho(d) = e^{-d/I} \tag{5.1}$$

where $\rho(d)$ is spatial correlation , d is the distance between two points and I is integral scale. (See Appendix B for the simulation of correlated variables and random fields.) The surface and river flow parameters were identical to those in the conceptual model. The key parameters in the reference model are presented in Table 5.1 and the generated random conductivity fields are illustrated in Fig. 5.2

Table 5.1: Key reference model parameters. The hydraulic conductivities are lognormal distributed, μ is mean value, σ is standard deviation and I is integral scale.

		μ	σ	I
Net precipitation [mm year^{-1}]		315	-	-
Conductivities	K_h	1.5E-04	7.5E-05	500
- upper aquifer [m s^{-1}]	K_v	1.0E-07	-	-
Conductivities	K_h	1.0E-08	-	-
- aquitard [m s^{-1}]	K_v	1.0E-08	5.0E-09	500
Conductivities	K_h	5.0E-04	2.5E-04	500
- lower aquifer [m s^{-1}]	K_v	1.0E-06	-	-

The model area was discretised into 470 elements in the horizontal plane and 5 layers in the vertical plane. A steady state flow simulation was performed using the FLEXFLOW model (see Chapter 2). The capture zones were derived from a forward particle tracking simulation with a particle density of 100 per km^2 in the horizontal plane. The groundwater heads obtained are presented in Fig. 5.3

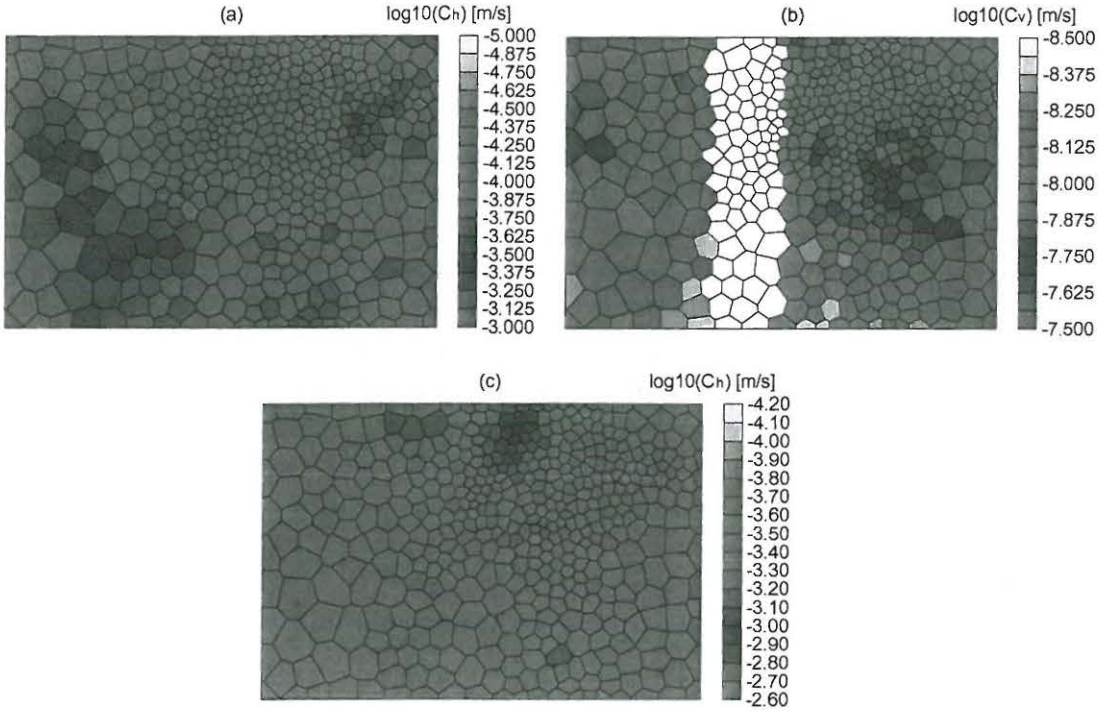


Figure 5.2: Random conductivity images. Horizontal conductivity in upper aquifer (a), vertical conductivity in aquitard (b), horizontal conductivity in lower aquifer(c). Note that the aquitard (b) is not continuously distributed in the study area: the sandy window belongs to the lower aquifer (geological model B).

Fig. 5.3 illustrates the fact that the equipotentials and therefore the flow direction were very different in the upper and lower aquifer. In the upper aquifer the equipotentials seemed to be controlled mainly by the river course, whereas in the lower aquifer they appeared to be controlled mainly by the presence of the abstraction well. As Fig. 5.4 shows, this complicates the flow pattern and influences the shape of the capture zone (Figs. 5.3 and 5.4).

5.3 Quantification of model error

The conceptual model incorporated model errors arising from the rough zonation of the hydraulic conductivities, and for this reason the model was not expected to represent small scale fluctuations in the hydraulic head. Gelhar (1986) has shown that the model error variance, s_h is a function of the variance of the log transformed hydraulic conductivities, $\sigma_{\ln K}$, the correlation length, α_l, and the hydraulic gradient, J:

$$s_h^2 = \frac{1}{3}\sigma_{\ln K}^2 \alpha_l^2 J^2 \tag{5.2}$$

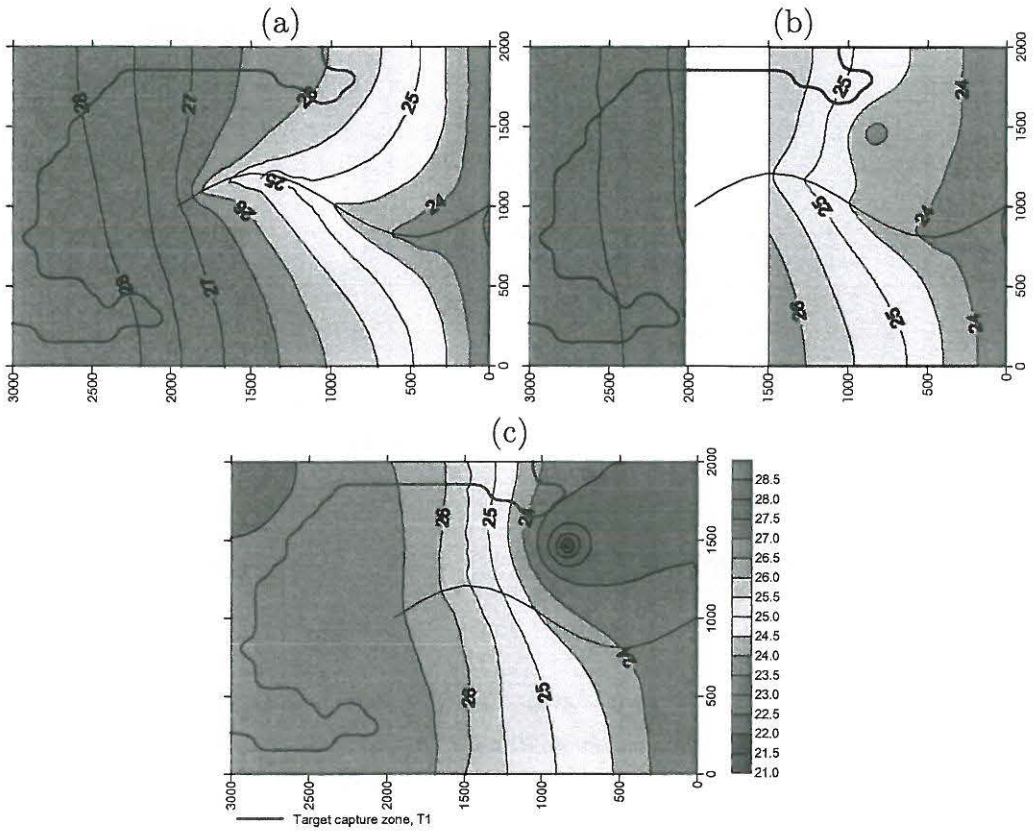


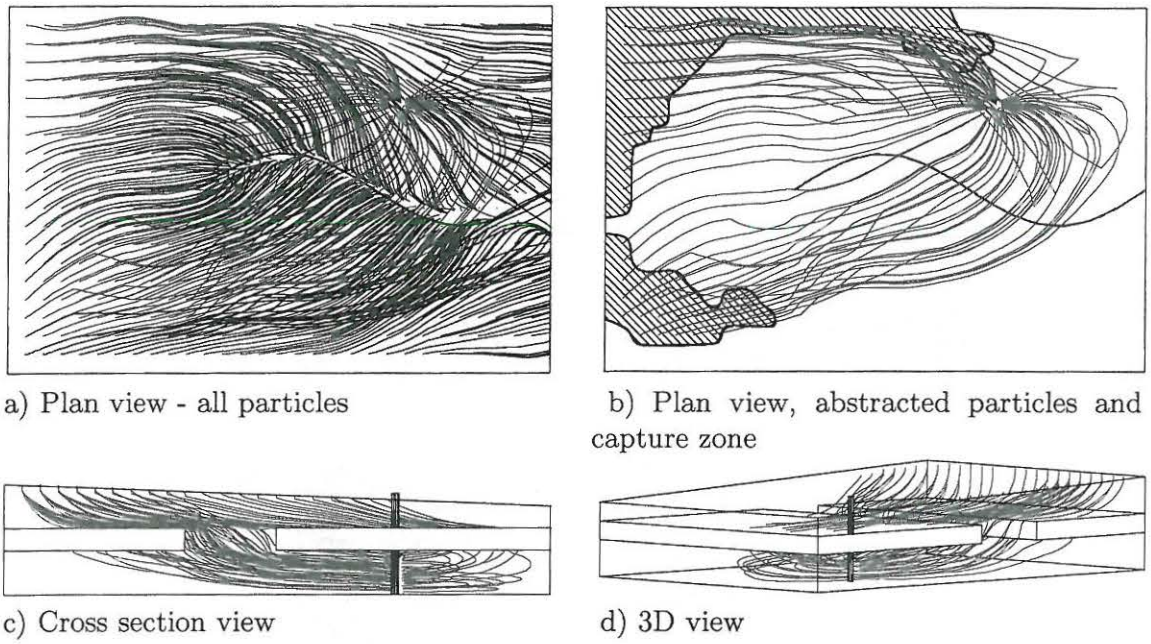
Figure 5.3: Groundwater heads in upper aquifer (a) aquitard (b) and lower aquifer (c).

In the present synthetic setup the correlation length, the standard deviation of the log transformed hydraulic conductivities and the hydraulic gradient were known, and the error arising from small scale heterogeneity, s_h, could be estimated, see Table 5.2.

The expected error, s_{RHS}, was also estimated by comparing nodal head values from the reference model with nodal head values from a model with homogeneous conductivities in the aquitard and the aquifers.

$$s^2_{RHS} = \frac{1}{N_n}\sum_{i=1}^{N_n}\left(\left(\frac{1}{N_n}\sum_{i=1}^{N_n}\left(\psi_i^r - \psi_i^h\right)\right) - \left(\psi_i^r - \psi_i^h\right)\right)^2 \tag{5.3}$$

N_n is the number of nodes, while ψ_i^r and ψ_i^h is the i^{th} nodal head value in the reference model and in the model with homogeneous conductivities respectively. The expected errors presented in Table 5.2 were used in the GLUE analysis used to design the likelihood functions applied.



a) Plan view - all particles

b) Plan view, abstracted particles and capture zone

c) Cross section view

d) 3D view

Figure 5.4: Flow path and capture zone. Particle density is 100 per km^2. Black particles leave the model area over the eastern boundary, blue particles leave the model area by the river and red particles are abstracted by the well

Table 5.2: Expected error due to small scale heterogeneity in the lower aquifer.

$\sigma_{\ln K}$ [m s^{-1}]	α_l [m]	J [-]	s_h [m]	s_{RHS} [m]
0.5	500	1.25	0.18	0.17

5.4 Monte Carlo simulations

From the conceptual model seven parameters were selected to be stochastic and 200,000 simulations were performed with the flow and transport models. Prior to each simulation a randomised setup (Monte Carlo method) of these seven parameters was performed on the basis of their distribution and range, as presented in Table 5.3.

A capture zone was delineated for every single simulation. The probability, $\hat{P}$ that a particle starting at point $(x_{0,i}, y_{0,i})$ is captured in the well is given as

$$\hat{P}\left(x_{0,l}, y_{0,l}\right) = \frac{\sum_{i=1}^{N_{sim}} L\left(\boldsymbol{\theta}_i\right) I_{\mathtt{CAP}_i}}{\sum_{i=1}^{K} L\left(\boldsymbol{\theta}_i\right)} \tag{5.4}$$

Table 5.3: Parameters used in Monte Carlo simulations. D, U and $L_{10}U$ respectively denote deterministic, uniform and Log_{10} uniform distribution.

Parameter		Distribution	Range
Geological model		U	P(A)=P(B)
Net precipitation [mm year^{-1}]		U	270 - 330
Conductivities	K_h	$L_{10}U$	1E-05 - 5E-03
- upper aquifer [m s^{-1}]	K_v	D	1E-07
Conductivities	K_h	$L_{10}U$	5E-9 - 5E-7
- aquitard [m/s]	K_v	$L_{10}U$	5E-9 - 5E-7
Conductivities	K_h	$L_{10}U$	5E-9 - 5E-7
- lower aquifer [m s^{-1}]	K_v	$L_{10}U$	5E-9 - 5E-7

where N_{sim} is the number of simulations and I_{CAP} is an indicator function that assigns the value 1 if $(x_{0,i}, y_{0,i})$ is within the i^{th} capture zone, CAP_i and 0 everywhere else:

$$I_{\text{CAP}_i} = \begin{cases} 1 & \text{if } (x_{0,i}, y_{0,i}) \in \text{CAP}_i \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

$L(\boldsymbol{\theta}_i)$ is the likelihood of the i^{th} simulation. All individual Monte Carlo simulations have, by definition, equal likelihood (weight), $L(\boldsymbol{\theta}_i) = \frac{1}{N_{obs}}$.

Based on the point probabilities, $\hat{P}(x_{0,l}, y_{0,l})$, a continuous probability surface was estimated by simple interpolation, Fig. 5.5.

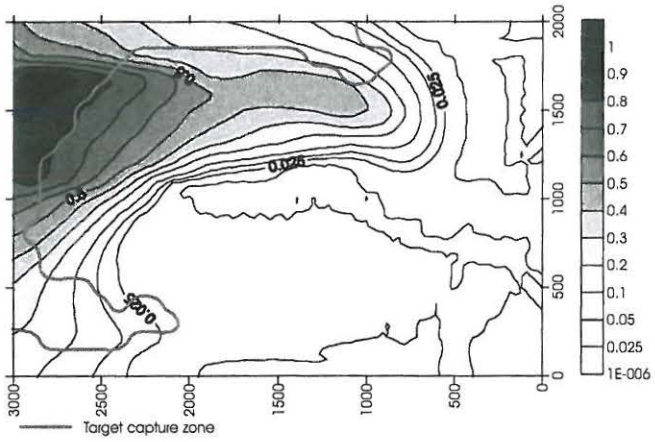


Figure 5.5: Target capture zone and Monte Carlo capture zone distribution.

The α% uncertainty bounds enclose the area characterised by $\frac{1-\alpha}{2} < \hat{P}(x_{0,l}, y_{0,l}) < \frac{1+\alpha}{2}$. The 95% uncertainty bounds thus enclose the area where $0.025 < \hat{P}(x_{0,l}, y_{0,l}) <$

0.975. The 95% prediction zone is therefore the zone with a probability greater than 2.5%

The 95 % capture zone region obtained from the Monte Carlo simulations did not cover the reference capture zone. This means that the quite wide parameter distribution used in the Monte Carlo simulation does not necessarily yield results that are safe within a reasonable prediction interval, as one might intuitively expect.

5.5 Observation data

Two observation data sets were extracted from the reference model. Data set **A** was used in the calibration and included seven head observations and one river discharge observation. The data set represents a typical situation where head observations are present only in the lower aquifer and only a few river discharge stations are present, Fig. 5.6. Data set **B** contained additional data used for validation, and included five head measurements and another river discharge measurement station, Fig. 5.6.

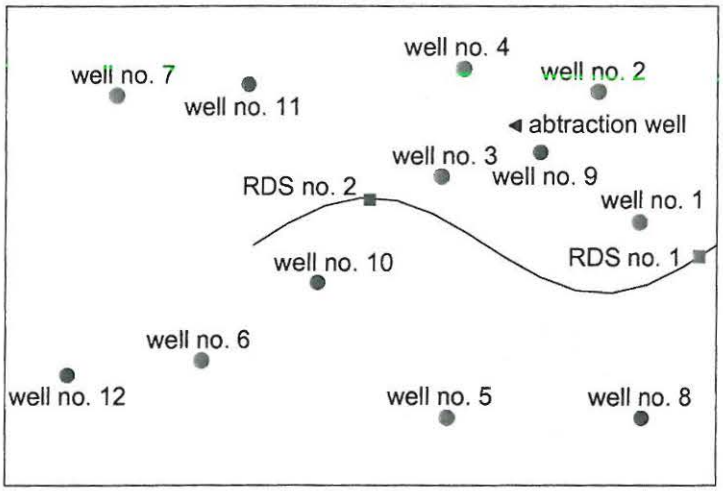


*Figure 5.6: Location of observation points, wells nos. 1-7 and river discharge station RDS no. 1 were used during calibration (red, data set **A**) and well no. 8-12 and RDS no. 2 were used as additional data in the validation procedure (blue, data set **B**)*

The head observations were affected only by model errors and the expected standard error therefore consisted only of the model error term, see Table 5.2. The standard error in the river discharge measurement was estimated at 10% of measured value from the general estimate presented in section 3.2.2.

5.6 GLUE analyses

The GLUE methodology and the likelihood measures referred to in this section are described in Chapter 4. The likelihood functions used in this study consisted of a point observation likelihood measure and an inference or combination function that linked the point observation likelihood measure to a global likelihood measure for the parameter set under consideration.

Based on the Monte Carlo simulations mentioned above, a number of GLUE analyses were performed. The first four analyses, sections 5.6.1, 5.6.2 and 5.6.3, dealt with the effect of including different types of observation data in the conditioning. The succeeding twelve analyses (2.1-2.4, 3.1-3.4, 4.1-4.4) concerned sensitivity with regard to the likelihood function, which was evaluated through the point likelihood function, (section 5.6.4), the inference function (section 5.6.5) and the rejection level (section 5.6.6). The reliability of any Monte Carlo-based method depends on the number of simulations; this issue is discussed in section 5.6.7. Finally the GLUE solution was validated against reference head and discharge data, see section 5.6.8.

The capture zone was estimated as described in section 5.4, with the exception that the single simulations now had an individual likelihood. The probability, $\hat{P}$ that a particle starting at point $(x_{0,i}, y_{0,i})$ is captured in the well is:

$$\hat{P}\left(x_{0,l}, y_{0,l}\right) = \frac{\sum_{i=1}^{N_s im} L\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}\right) I_{\mathrm{CAP}_i}}{\sum_{i=1}^{K} L\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}\right)} \tag{5.6}$$

where $L\left(\boldsymbol{\theta}_i|\boldsymbol{\psi}\right)$ is the conditional likelihood of the i^{th} simulation.

5.6.1 Conditioning based on head data

Figure 5.7 illustrates the results from the GLUE analysis, where the Monte Carlo simulations were conditioned on seven head observations in the lower aquifer, Fig. 5.6 - dataset **A**.

A point likelihood was calculated for each of the seven observations wells, using the Gaussian likelihood function:

$$L\left(\psi_j^*|\boldsymbol{\theta}\right) = \frac{1}{\sqrt{2\pi}\sigma_h} e^{-\left(\frac{\left(\psi_j^* - \psi_j(\boldsymbol{\theta})\right)^2}{2{\sigma_h}^2}\right)} \tag{5.7}$$

and likelihoods were combined by the geometric mean inference function:

$$L\left(\boldsymbol{\psi}^*|\boldsymbol{\theta}\right) = \sqrt[7]{\prod_{j=1}^{7} L\left(\psi_j^*|\boldsymbol{\theta}\right)} \tag{5.8}$$

Based on the considerations in section 5.3, the expected head error, σ_h was set to 0.2 m in all observations wells. The results are shown on Fig. 5.7. Of the 200,000 simulations, 2,053 simulations were accepted, N_{acc}. The capture zone estimate, Fig. 5.7b, shows little predictive capability - the 95% probability zone does not cover the reference capture zone (RCZ) and its shape is not recognisable from the GLUE results.

In Fig. 5.7c the accumulated likelihood for each of the two geological models is presented and they are approximately equal. From this analysis it is not possible to distinguish between the two geological models.

The so-called "scatter plot" (parameter value plotted against likelihood for all accepted simulations) is presented for the remaining six parameters, Figs. 5.7d - 5.7i. The scatter plots for net precipitation, horizontal conductivity in the aquitard and vertical conductivity in the lower aquifer show identical or close to identical solutions in the entire parameter range sampled. This may be caused by equifinality problems: given the current level of observations (and the reliability of these) it is not possible to constrain the solutions. Note that the plot range corresponds to the sampling range.

On the other hand the intervals of horizontal conductivity in the upper and lower aquifer and of vertical conductivity in the aquitard are narrowed by the GLUE analysis.

5.6.2 Conditioning based on river discharge data

The river discharge observation near the outlet (RDS no. 1) of the catchment was used in the conditioning, Fig. 5.8.

The likelihood for the river discharge station was calculated using the Gaussian likelihood function:

$$L\left(\psi^*|\boldsymbol{\theta}\right) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\left(\frac{\left(\psi^* - \psi(\boldsymbol{\theta})\right)^2}{2{\sigma_r}^2}\right)} \tag{5.9}$$

where σ_r is the expected standard error in the discharge observation. On the basis of section 3.2.2, σ_r was estimated at 10% of the observed discharge.

The likelihood measure is quite unrestrictive and 23,720 - close to 12% - of all simulations were accepted.

The results showed that the 95 % probability zone covers the reference capture zone, although the high probability zone is located in the lower left corner and in the upper part of the catchment, Fig. 5.8b. The river discharge observation seems to have had a considerable influence on the intervals of the horizontal conductivity in the upper aquifer, Fig. 5.8e and some influence on the horizontal conductivity in the lower aquifer, Fig. 5.8i. The lower interval of the net precipitation tended to have less likelihood than the upper interval. The accumulated likelihoods of the geological models are approximately equal, Fig. 5.8c. The remaining parameters seemed to be unaffected of the river discharge observation.

5.6.3 Conditioning based on head data and river discharge data

In the following two analyses the Monte Carlo simulations were conditioned on the basis of all observation in data set A: seven head observations in the lower aquifer and the river discharge station near the outlet, Fig. 5.6. The influence of river discharge data was gradually included. At first a rather high expected river discharge error of 20% was assumed, analysis 1.3 - Fig. 5.9. Later, the expected river discharge error was reduced to a more realistic level of 10%, analysis 1.4 - Fig. 5.10.

During analysis 1.3 245 simulations were accepted, while as the influence of river discharge data was increased (analysis 1.4) only 26 simulations were accepted. Analysis 1.4 appears to yield a better estimate for the capture zone - the 95% prediction zone covers the RCZ and the overall shape of the probability contours becomes very similar to the contours of the RCZ. When the accumulated likelihood of the geological models was considered it was very interesting to observe that the model failed to predict the geological model B: 93% to 99% of all likelihood was on model A, Figs. 5.9c and 5.10c. Another interesting result was that the simulation yielding the highest likelihood was based on a parameter set containing the geological model B, and this parameter solution is actually comparable with the solution that will be obtained if a classical regression method is applied to the problem.

In the range of accepted parameter values for horizontal conductivity in the upper and lower aquifer and for vertical conductivity in the aquitard, a biased reduction was also observed in comparison to analyses 1.1 and 1.2, Figs. 5.7 and 5.8.

The net precipitation, horizontal conductivity in the aquitard and vertical conductivity in the lower aquifer still seemed to be unaffected by the observations. However, there was a tendency towards a higher likelihood for the upper range

of the net precipitation interval in analysis 1.4. Note that there may be quite large statistical uncertainty in analysis 1.4, due to the relatively low number of accepted simulations.

5.6.4 Influence of point likelihood function

The influence of the choice of point likelihood function was examined. Four different likelihood functions were applied: the Gaussian, uniform, trapezoidal and triangular likelihood functions, see section 4.1.4. A rejection level of three times the expected standard error, together with the geometric mean inference function, was used for all four analyses; see Table 5.4 for details. In all analyses the conditioning was based on data set A (seven head observations and one river discharge observation).

Table 5.4: GLUE results - influence of the choice of likelihood function

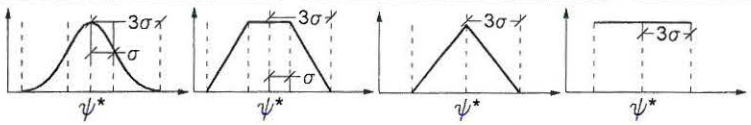
Analysis:	2.1	2.2	2.3	2.4
Likelihood function	 σ_h=0.2m σ_r=10%	σ_h=0.2m σ_r=10%	σ_h=0.2m σ_r=10%	σ_h=0.2m σ_r=10%
Inference function	geo. mean	geo. mean	geo. mean	geo. mean
N_{obs}	8	8	8	8
N_{acc}	26	26	26	26
95% pred. zone	2.14	2.23	2.22	2.33

Fig. 5.11 illustrates the capture zone distributions for the different point likelihood functions.

The choice of likelihood function proved to make no significant difference in the capture zone distribution. The 95% prediction zone tended to increase as the likelihood function became more voluminous (e.g. from Gaussian to uniform), Table 5.4. The area increased by approximately 10% from the Gaussian to the uniform likelihood function.

The influence of the point likelihood measure may depend on the inference function. The product inference function will show greater differences between the different likelihood functions, while the arithmetic mean will show smaller differences between them.

Calibration - conditioning based on head data

Analysis:	Likelihood function :	Gaussian	N_{obs}	:	7
1.1		σ_h=0.2m, σ_r=10%	N_{acc}	:	2053
	Inference function :	geometric mean	95% pred. zone	:	2.14 km^2

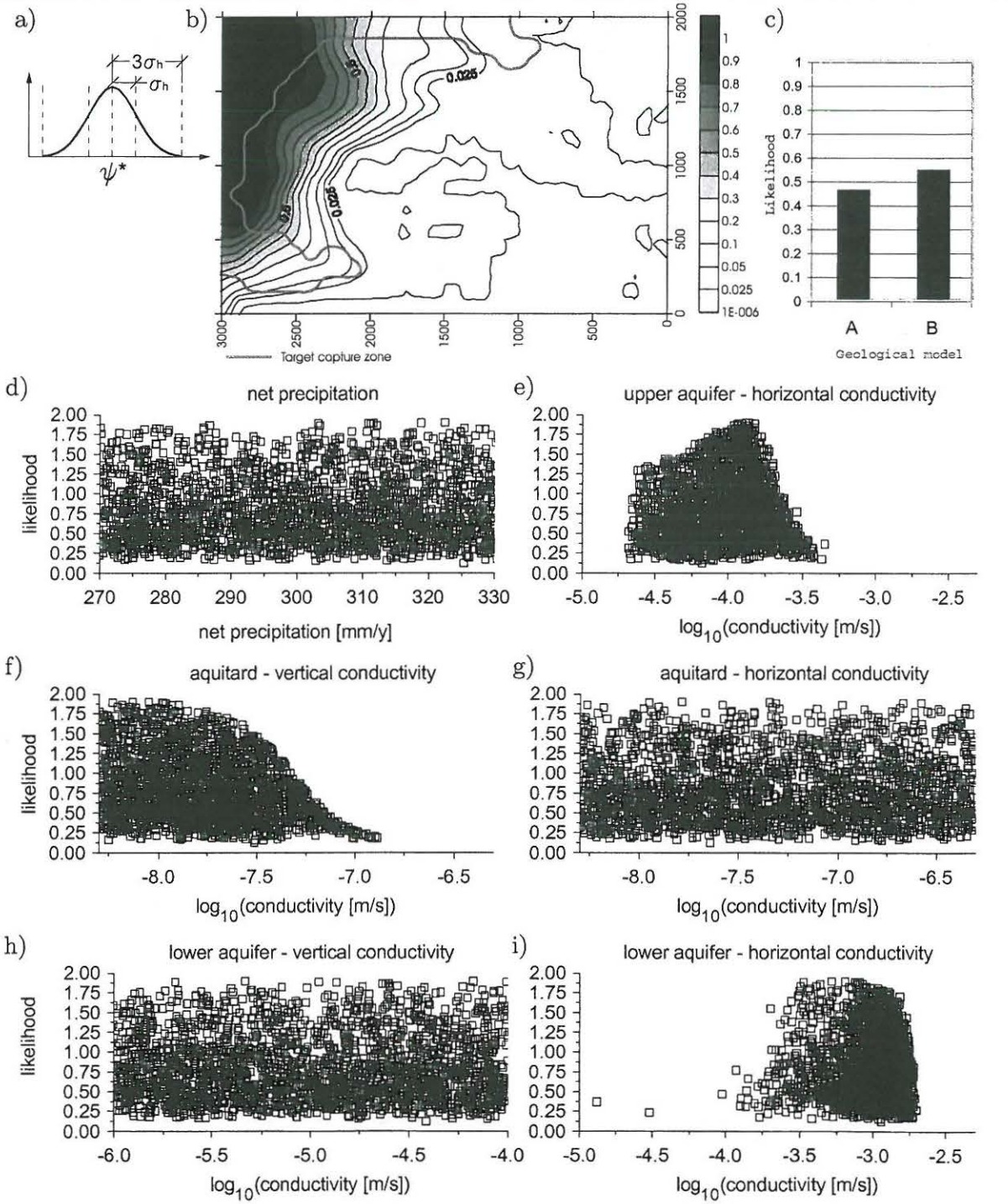


Figure 5.7: GLUE results - conditioning based on head observations: Applied likelihood function (a), capture zone distribution (b), accumulated likelihood of geological model A and B (c), and parameter likelihood scatter plot (d-i).Note that the plot range corresponds to the sampling range. See table at top of page for likelihood function, inference function, no. of observations used in the conditioning, N_{obs}, no. of accepted simulations, N_{acc} and area of the 95% prediction zone.

Calibration - conditioning based on river discharge data

Analysis:	Likelihood function :	Gaussian	N_{obs}	:	1
1.2		σ_r=10%	N_{acc}	:	23720
	Inference function :	geometric mean	95%*pred.zone*	:	2.14 km^2

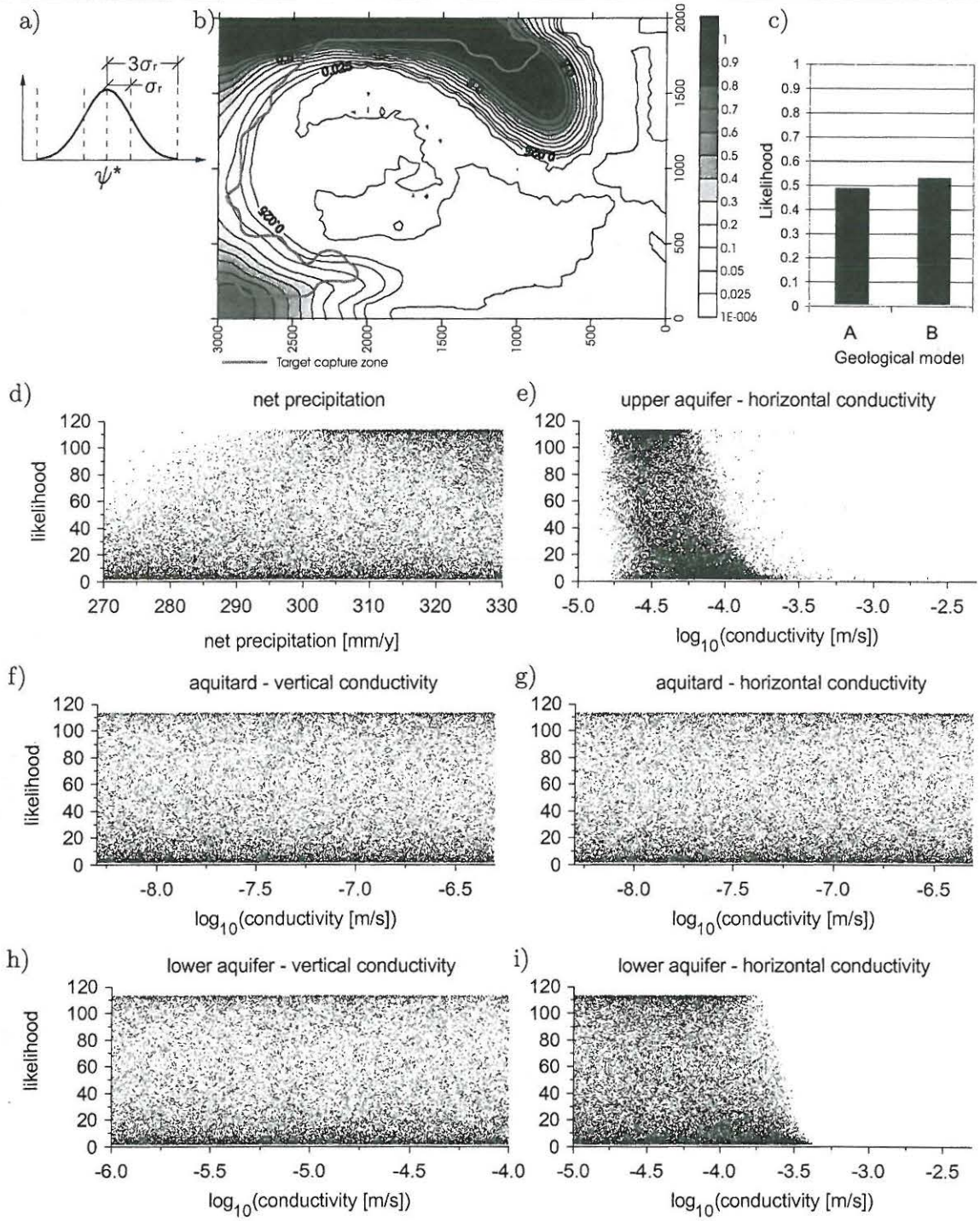


Figure 5.8: GLUE results - conditioning based on river discharge observations: Applied likelihood function (a), capture zone distribution (b), accumulated likelihood of geological model A and B (c), and parameter likelihood scatter plot (d-i).Note that the plot range corresponds to the sampling range. See table at top of page for likelihood function, inference function, no. of observations used in the conditioning, N_{obs}, no. of accepted simulations, N_{acc} and area of the 95% prediction zone.

Calibration - conditioning based on head and river data - large discharge error

Analysis:	Likelihood function :	Gaussian	N_{obs}	:	8
1.3		σ_h=0.2m, σ_r=20%	N_{acc}	:	245
	Inference function :	geometric mean	95%*pred.zone*	:	2.02 km^2

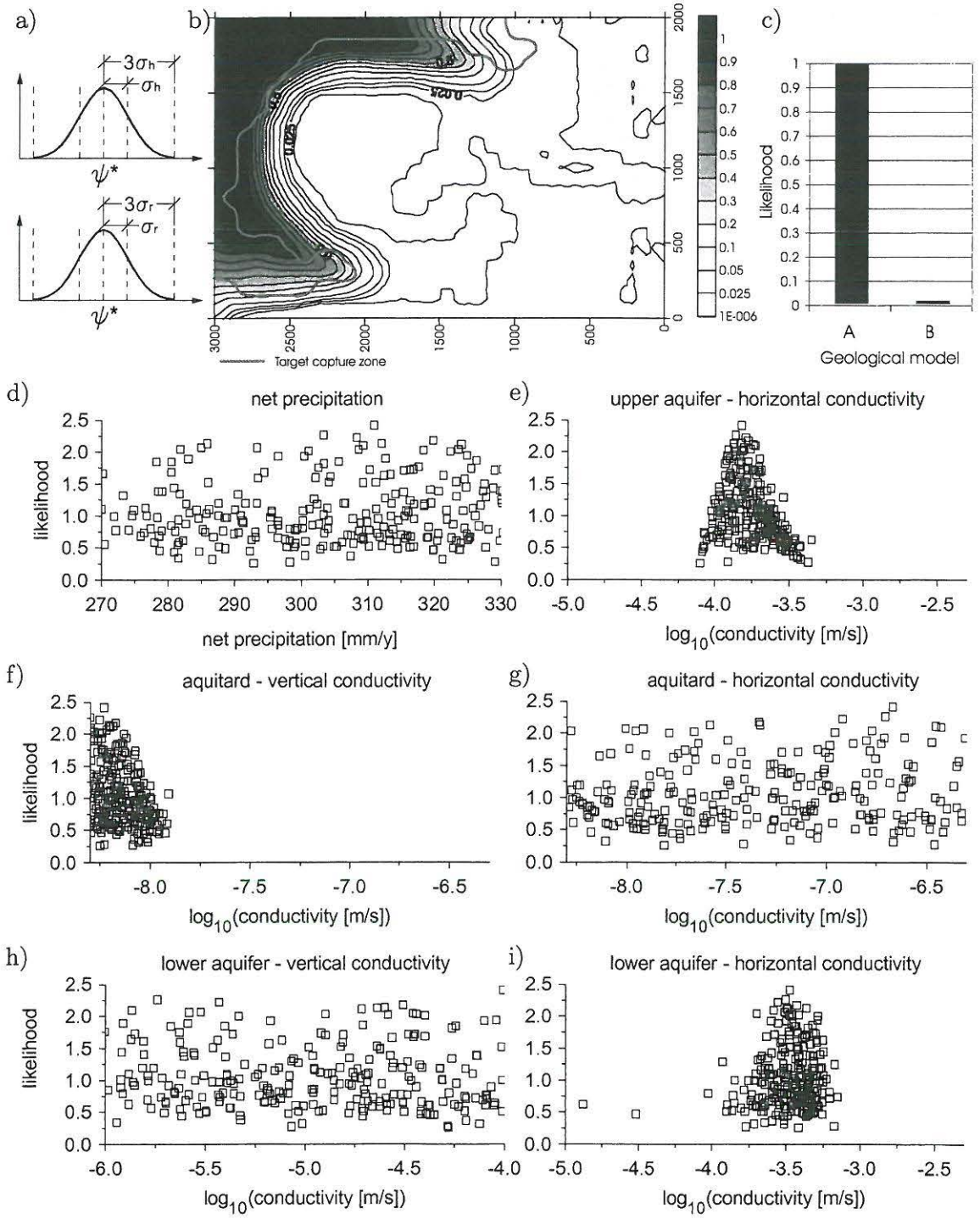


Figure 5.9: GLUE results - conditioning based on head and river discharge observations: Applied likelihood function (a), capture zone distribution (b), accumulated likelihood of geological model A and B (c), and parameter likelihood scatter plot (d-i).Note that the plot range corresponds to the sampling range. See table at top of page for likelihood function, inference function, no. of observations used in the conditioning, N_{obs}, no. of accepted simulations, N_{acc} and area of the 95% prediction zone.

Calibration - conditioning based on head and river data - small discharge error

Analysis:	Likelihood function :	Gaussian	N_{obs}	:	8
1.4		σ_h=0.2m, σ_r=10%	N_{acc}	:	26
	Inference function :	geometric mean	95%*pred.zone* :		2.14 km^2

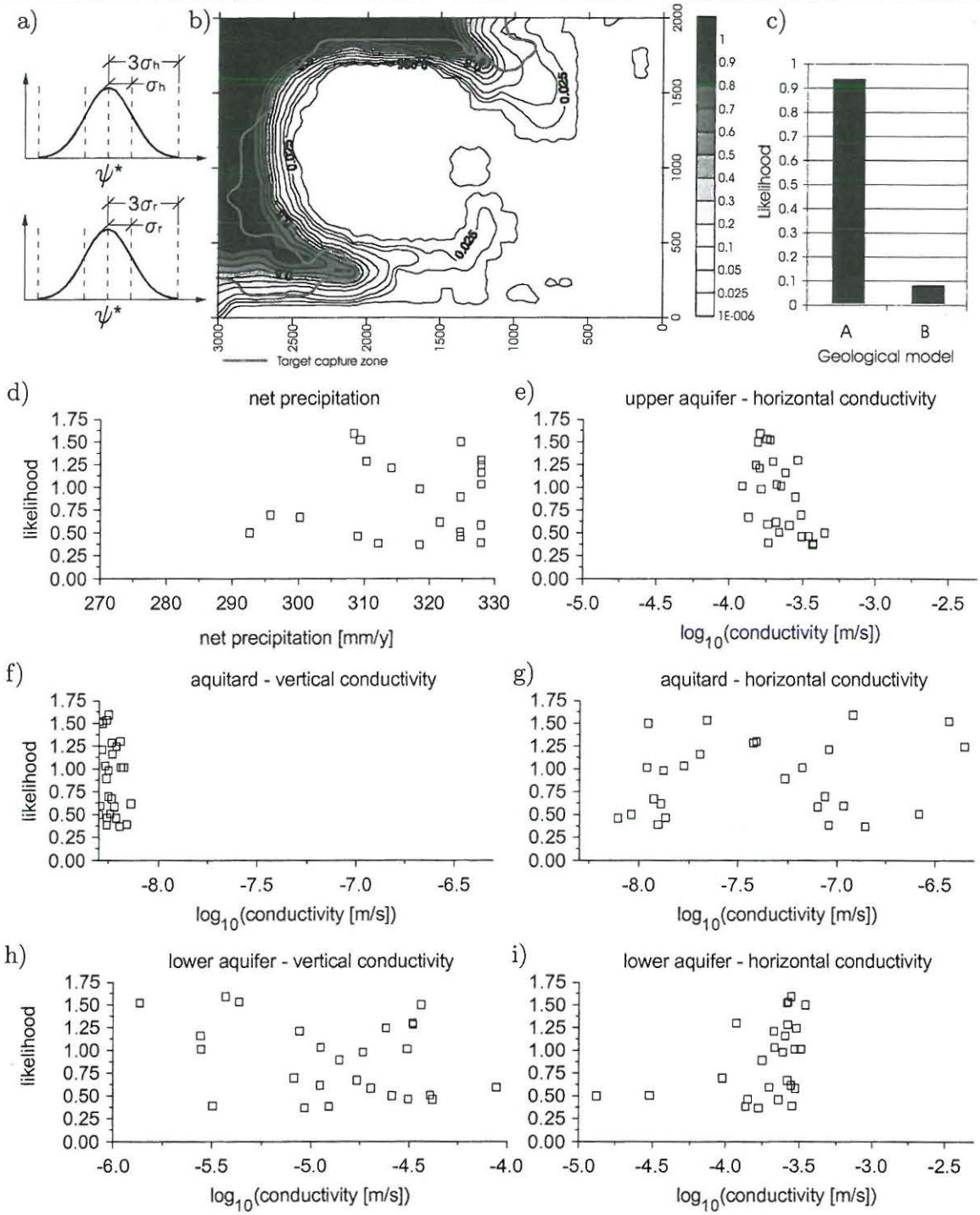


Figure 5.10: GLUE results - conditioning based on head and river discharge observations: Applied likelihood function (a), capture zone distribution (b), accumulated likelihood of geological model A and B (c), and parameter likelihood scatter plot (d-i).Note that the plot range corresponds to the sampling range. See table at top of page for likelihood function, inference function, no. of observations used in the conditioning, N_{obs}, no. of accepted simulations, N_{acc} and area of the 95% prediction zone

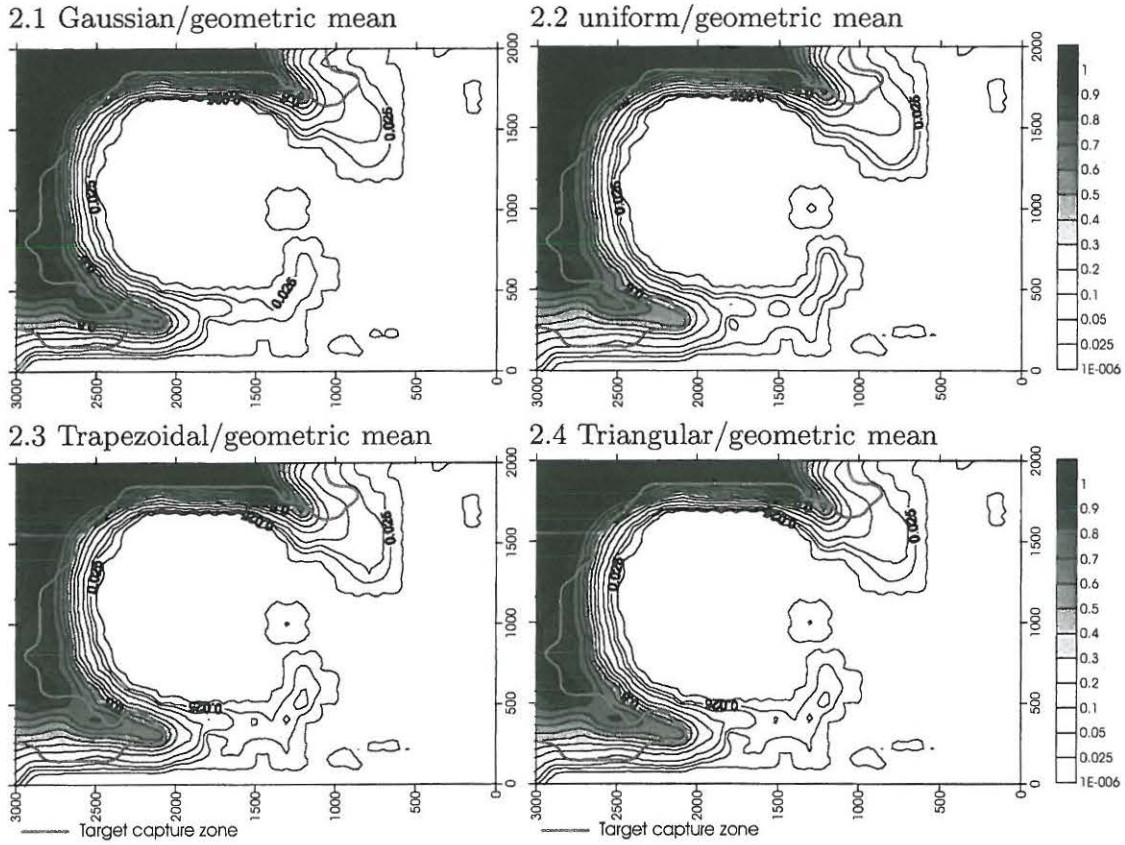


Figure 5.11: Capture zone distributions with four different point likelihood functions. The Geometric mean inference function was used in all estimates

5.6.5 Influence of inference function

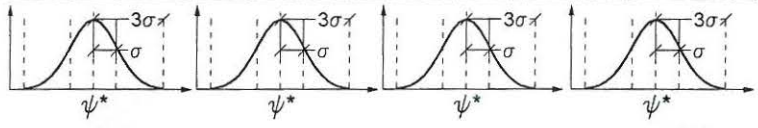
The influence of the choice of inference function was examined. The geometric mean, product, max. and arithmetic mean inference function were tested using Gaussian likelihood in all analyses, Table 5.5. Note that the max. and the arithmetic mean inference function are less restrictive, and consequently a large number of simulations were accepted, Table 5.5.

The capture zone distributions are presented in Fig. 5.12.

The capture zone distribution from the Gaussian point likelihood function, using product inference, Fig. 5.12b, are similar to the distribution found from Monte Carlo simulations based on the statistics derived from a statistical framework in which non-linearity is accounted for.

In addition to a small region in the upper right corners, the 95% prediction zone covered the reference capture zone. If the number of observations increases it is expected that the product inference function will underestimate the capture zone distribution, since the likelihood function will become steeper.

Table 5.5: GLUE results - influence of inference function.

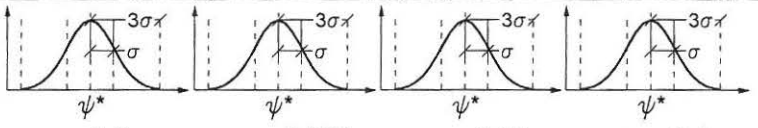
Analysis:	3.1	3.2	3.3	3.4
Likelihood function				
	σ_h=0.2m σ_r=10%	σ_h=0.2m σ_r=10%	σ_h=0.2m σ_r=10%	σ_h=0.2m σ_r=10%
Inference function	geo. mean	product	max	arith. mean
N_{obs}	8	8	8	8
N_{acc}	26	26	130366	121469
95% pred. zone	2.14	1.69	2.89	2.94

The geometric mean inference function predicts a capture zone distribution that is very similar in shape to the reference zone, while the shape of the capture zone distribution obtained from the max. and arithmetic mean inference deviates from the reference capture zone. The areas corresponding to the 95% prediction zones are presented in Table 5.5.

5.6.6 Influence of rejection level

The influence of the rejection level was examined by increasing the expected standard error of the observations from 0% to 100%. The Gaussian point likelihood function was used, together with the geometric mean inference function; see parameters in Table 5.6.

Table 5.6: GLUE results - influence of rejection level.

Analysis:	4.1	4.2	4.3	4.4
Likelihood function				
	σ_h=0.2m σ_r=10%	σ_h=0.225m σ_r=12.5%	σ_h=0.25m σ_r=15%	σ_h=0.4m σ_r=20%
Inference function	geo. mean	geo. mean	geo. mean	geo. mean
N_{obs}	8	8	8	8
N_{acc}	26	68	138	2167
95% pred. zone	2.14	2.10	2.14	2.49

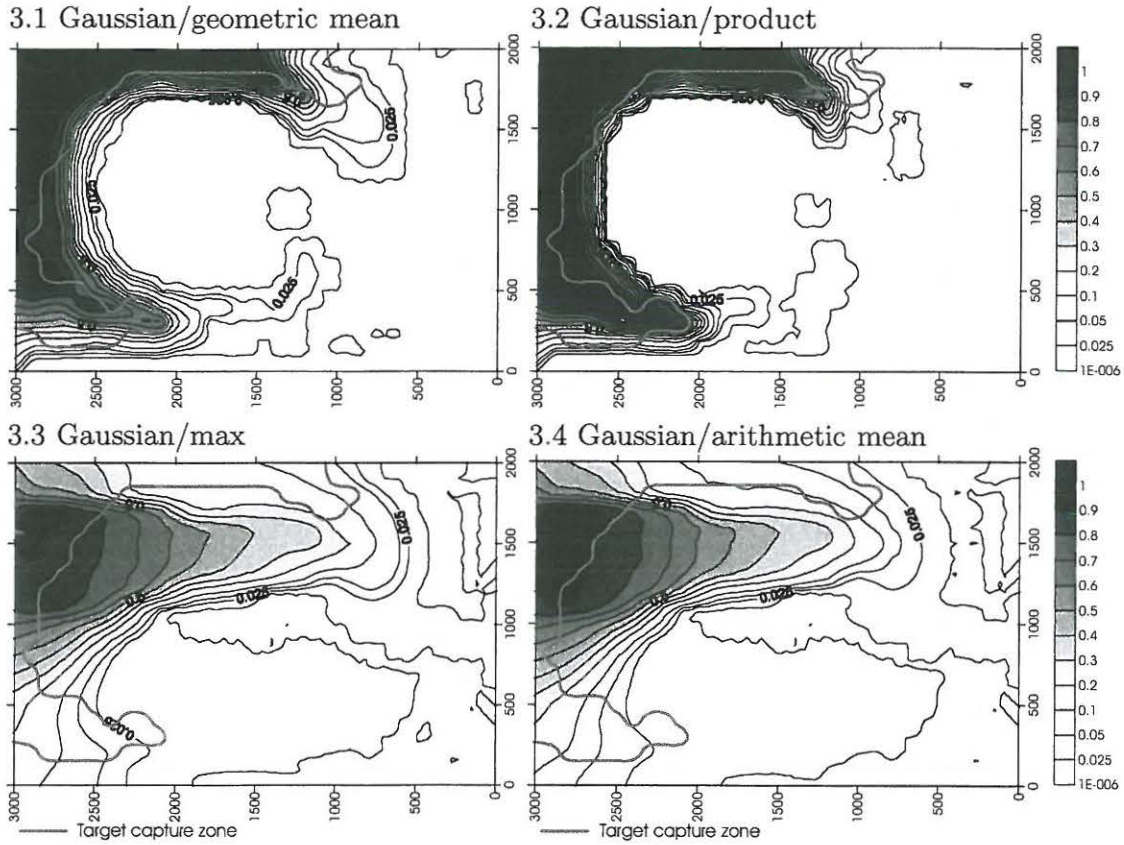


Figure 5.12: Capture zone distributions with four inference functions. The Gaussian point likelihood function was used in all estimates

The choice of rejection level does not have any significant influence up to at least a 25% enhancement of the expected errors (rejection level), whereas a doubling of the expected errors results in a capture zone distribution that has changed dramatically compared to the starting point: analysis 4.1, Table 5.6 and Fig. 5.13.

5.6.7 The necessary number of simulations

The reliability of capture zone distribution estimates depends on the number of accepted simulations. The degree of error arising from a finite number of simulations is not given beforehand. However, it can be approximated if the weight of the accepted simulations is approximately equal, as in the Monte Carlo simulations. The standard error $s(x_{0,l}, y_{0,l})$ of the capture zone probability, $P(x_{0,l}, y_{0,l})$, in point $(x_{0,l}, y_{0,l})$ is given by (Rubinstein 1981)

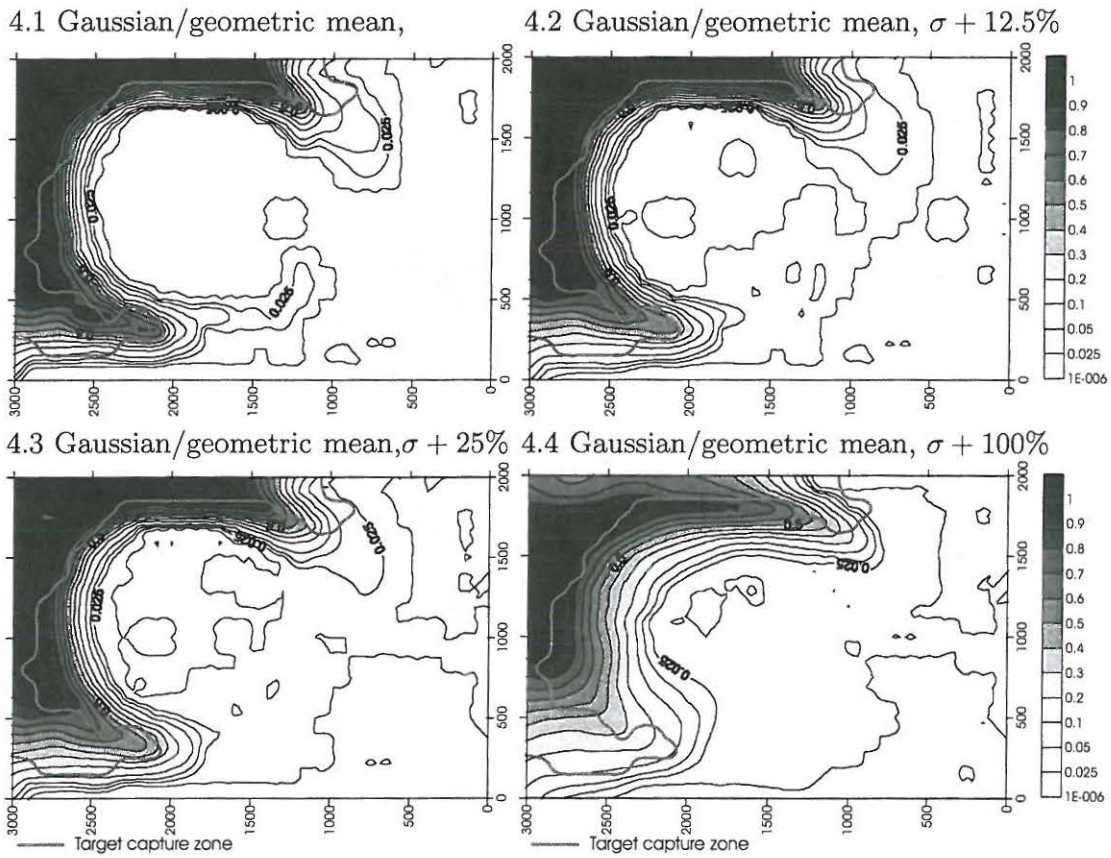


Figure 5.13: Influence of rejection level. Expected error increased by 0, 12.5, 25 and 100%

$$s\left(x_{0,l}, y_{0,l}\right) = \sqrt{\frac{P\left(x_{0,l}, y_{0,l}\right)\left[1 - P\left(x_{0,l}, y_{0,l}\right)\right]}{N_{acc}}} \tag{5.10}$$

where $P\left(x_{0,l}, y_{0,l}\right)$ is the unknown true point probability and N_{acc} is the number of accepted simulations.

From Eq. 5.6 the point probability can be estimated, and if we assume that the error is Gaussian the upper bound 95% confidence interval can be formulated as the estimate, $\hat{P}\left(x_{0,l}, y_{0,l}\right)$, plus two times the standard error $s\left(x_{0,l}, y_{0,l}\right)$:

$$P\left(x_{0,l}, y_{0,l}\right) = \hat{P}\left(x_{0,l}, y_{0,l}\right) + 2\sqrt{\frac{P\left(x_{0,l}, y_{0,l}\right)\left[1 - P\left(x_{0,l}, y_{0,l}\right)\right]}{N_{acc}}} \tag{5.11}$$

Solving Eq. 5.11 with $\hat{P}\left(x_{0,l}, y_{0,l}\right) = 0$ we can obtain the 95% confidence bound for the estimate of $P\left(x_{0,l}, y_{0,l}\right)$ as a function of the number of accepted simulations: that is, the upper 95% confidence interval given zero estimated probability.

In Fig 5.14, Eq. 5.11 with $\hat{P}\left(x_{0,l}, y_{0,l}\right) = 0$ is plotted as a function of N_{acc}. The graph illustrates the minimum safety level that can be considered given a specific

number of accepted simulations. E.g. if the number of accepted simulations is 150, the minimum considered safety level is approximately 2.5%

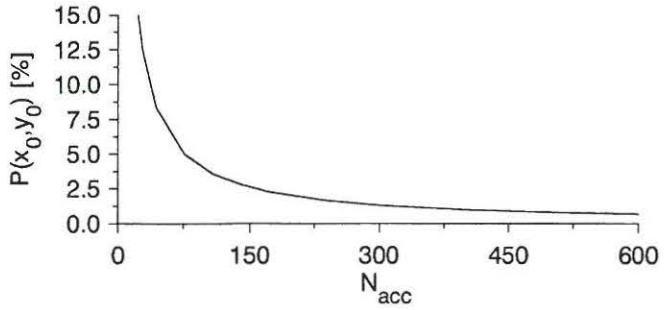


Figure 5.14: Minimum point probability as a function of accepted simulations - from Eq. 5.11, with $\hat{P}(x_{0,l}, y_{0,l}) = 0$

Solving Eq. 5.11 for any $\hat{P}(x_{0,l}, y_{0,l})$ will yield 95% $(\hat{P} + 2s(x_{0,l}, y_{0,l}))$ upper bound for the point probability. It may be possible by this means to account for the uncertainty due to a finite number of accepted simulations.

The number of accepted simulations is a function of the rejection level and the choice of inference function. The Gaussian likelihood function, using the geometric mean inference function, results in the relationship between accepted simulations and rejection level that is presented in Fig. 5.15.

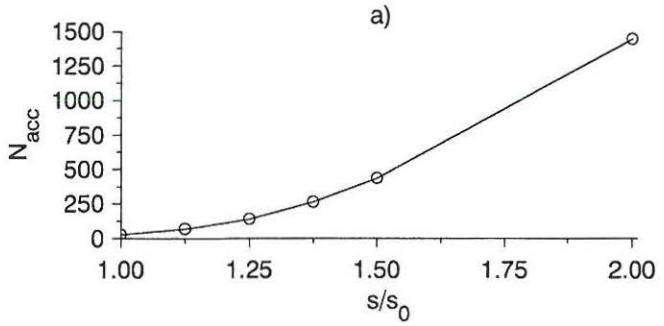


Figure 5.15: Number of accepted simulations as a function of rejection level (standard error). s_0 *corresponds to standard head deviation of 0.2 m and standard discharge deviation of 10%: data points found from numerical experiment.*

For this specific model setup Figs. 5.14 and 5.15 can now be used to obtain the minimum safety level as a function of the rejection level. E.g. a rejection level of $1.25\sigma_0$ yields about 140 accepted simulations, and 140 simulations yields a minimum safety level of 3%.

5.6.8 Validation of simulated heads

Five additional head observation points (wells nos. 8 - 12) and one additional river discharge station (RDS no.2) were included in the setup, Fig. 5.6 - data set B. Based on the GLUE analysis 1.4 (Gaussian likelihood function ($\sigma_h = 0.2m, \sigma_r = 10\%$) and geometric mean inference function) a likelihood distribution was found in all observations points (data set A and B). The likelihood distribution and the corresponding observation for heads and for river discharge are presented in Fig. 5.16.

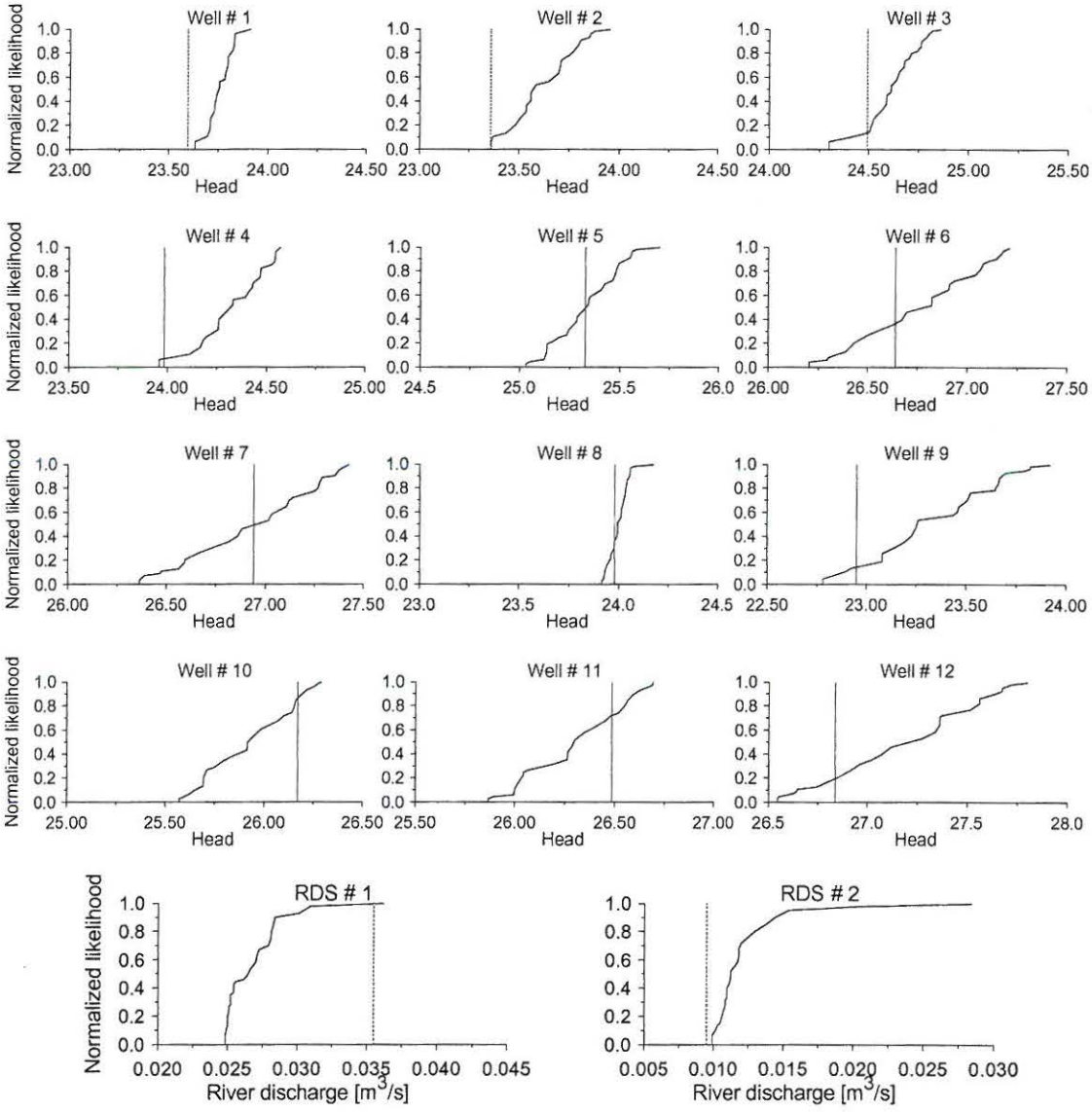


Figure 5.16: Head and river discharge likelihood distributions, $\sigma_h = 0.2, \sigma_r = 10\%$. The vertical line represents the reference value (observation)

The capacity to predict heads tended to decrease in the region close to the

abstraction well (wells nos. 1,2,3,4 and 9). Moreover the likelihood distribution for the river discharge showed weak predictive capability, and it should be noted that the observed river discharge in station 1 was close to the upper bound of the interval, while the river discharge in station 2 was below the lower bound. This indicates a rather significant difference between the reference model and the conceptual model with respect to the distribution of the lateral inflow to the river.

In three observations points (wells nos. 1,2 and station no. 2) the observed value was outside the simulated interval. This might indicate that the rejection level was too narrow.

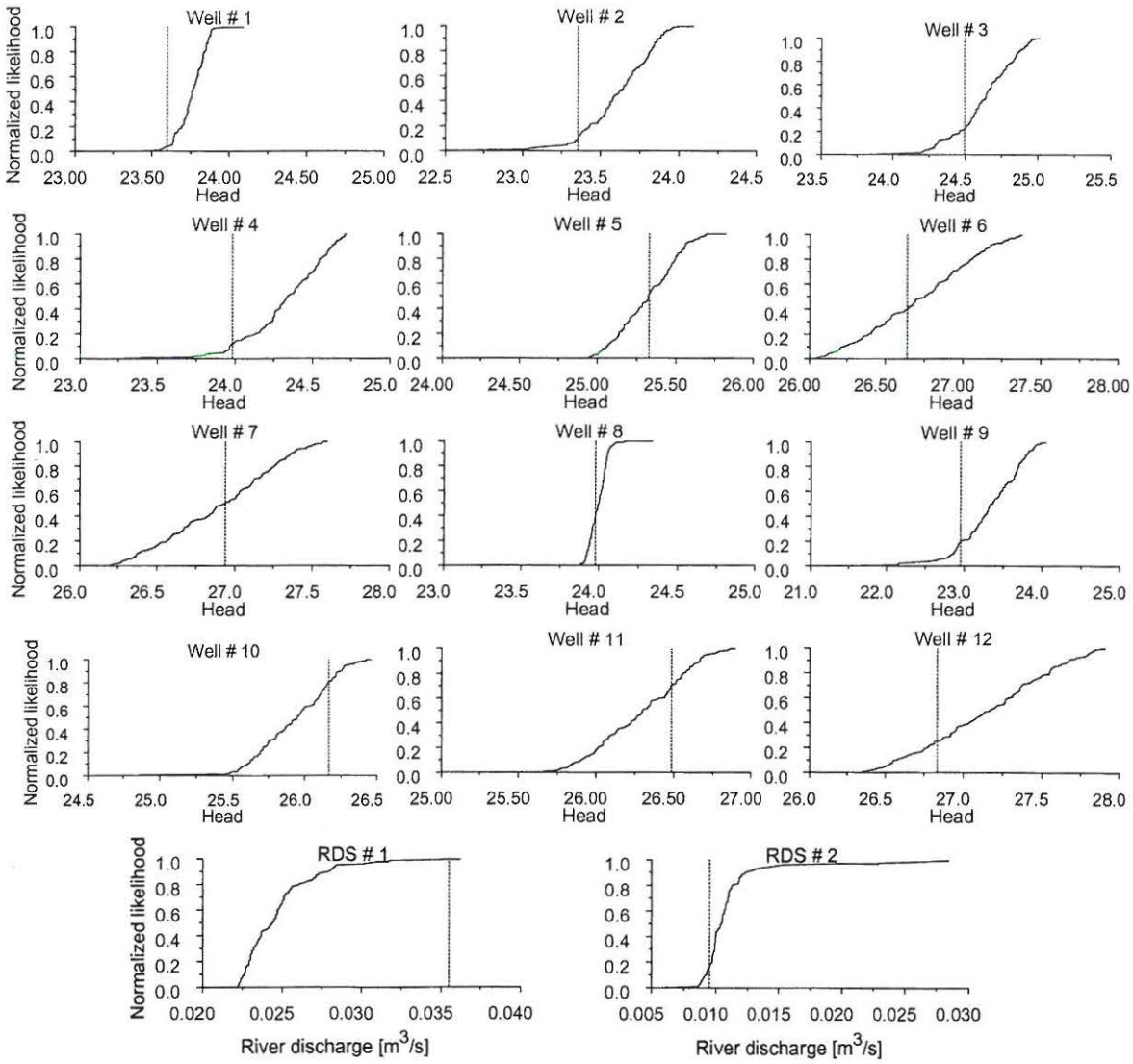


Figure 5.17: Head and river discharge likelihood distributions, $\sigma_h = 0.25$, $\sigma_r = 12.5\%$. The vertical line represents the reference value (observation)

Fig. 5.17 illustrates the likelihood distributions when the expected head and river discharge error is increased by 25% (analysis 4.3). The corresponding capture zone is found in Fig. 5.13c.

All observed values are now placed well within the simulated likelihood intervals.

5.7 Summary

The GLUE methodology was used to predict the capture zone distribution in a synthetic setup. A reference model was constructed from which a reference capture zone and corresponding head and river discharge "observations" were extracted. A conceptual model was formulated with seven unknown parameters. The hydraulic conductivities were modelled as random fields in the reference model and as homogeneous fields in the conceptual model. This deviation in the description was designed for the purpose of introducing models errors into the conceptual model.

A number of different likelihood measures were applied and the conclusions were:

- The type of point likelihood function has no significant influence.
- The capture zone estimate depends to a considerable degree on the choice of inference function. The max. and the arithmetic mean inference function yielded capture zone estimates that were very unlike the reference capture zone, whereas the product and geometric mean inference function produced capture zone estimates that were close to the reference capture zone.
- An increase in the rejection level of up to 25% does not significantly influence the capture zone estimates. A doubling of the rejection level will however change the estimate dramatically.
- It was not possible to predict the correct geological model on the basis of the GLUE analysis, nor was the most likely simulation based on the correct geological model.

The number of accepted simulations necessary in order to achieve a reliable estimate was discussed. If it is assumed that the likelihood values of the accepted simulations are approximately equal (which might be a rough approximation in relation to some likelihood functions) then the standard error of the likelihood estimate can be found. It was shown that approximately 150 accepted simulations were needed in order to make any predictions within a confidence level of 95%

Finally the likelihood distributions were found for twelve head observations points and two river discharge stations. The validation was based on Gaussian point likelihood functions and the geometric mean inference function. The expected errors in the observations were 0.2m for head and 10% for river discharge observations. These values correspond to "the best prior guess" based on information concerning geological heterogeneity, section 5.3, and the general expected uncertainty of river discharge, section 3.2.2. The validation showed that four observations were outside the predicted intervals. This indicates an over-restrictive likelihood measure. The expected errors were consequently increased by 25% and a new validation was performed. All observations were now placed well within the prediction intervals.

In this chapter it has been shown that the GLUE methodology is very sensitive to the choice of likelihood measure. The selection of a suitable likelihood measure may be an easy task in the case of a synthetic setup because the solution is known, whereas in real case studies this selection procedure may be difficult to carry out. The results show however that the suggested guidelines for selecting likelihood measures, as presented in section 4.2, seem to yield satisfactory results.

CHAPTER 6

Case B: Calibration and uncertainty estimation in a regional aquifer system

In the present chapter a GLUE analysis was applied to a Danish regional aquifer system covering the Gjern river system. The model setup was identical to that presented in Christensen et al. (1998). Christensen et al. (1998) estimated model parameters and their 95% confidence intervals by using a non-linear regression. Their 1998 study estimated stream flows, including linear confidence and prediction intervals, while in Christensen and Cooley (1999) linear and non-linear predictions intervals for head and stream flows were presented.

The aim of the present study was to apply the GLUE methodology to exactly the same problem (and numerical code) and to compare the prediction intervals obtained with those obtained by Christensen and Cooley (1999). The setup will be briefly described in this chapter; readers interested in further details are referred to Christensen et al. (1998).

6.1 Conceptual model

The Gjern catchment covers an area of 114 km^2 in eastern Jutland. The topographic water divide, the groundwater divide, and the location of streams are presented in Fig. 6.1

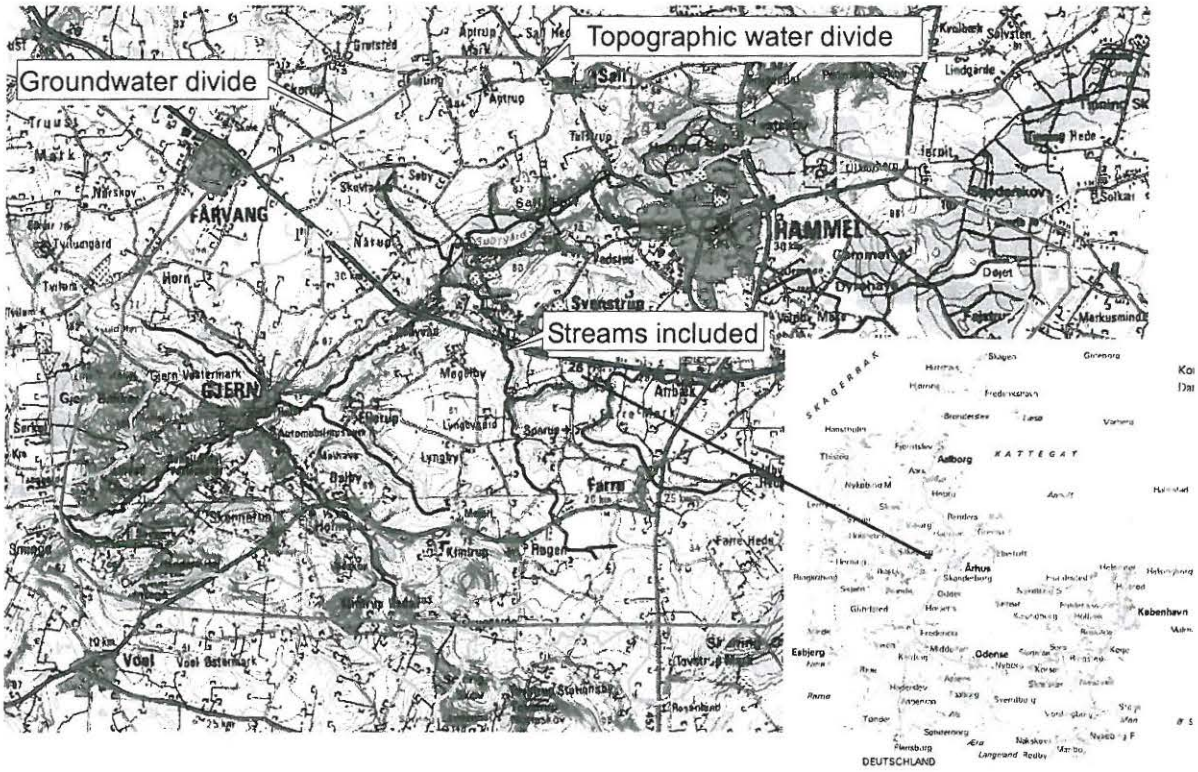


Figure 6.1: Groundwater divide, topographical water divide and location of streams

6.1.1 Geology

The main aquifer consisting of sand and gravel extends throughout the entire catchment, apart from a region in the southwest catchment corner. This main aquifer was divided into eight hydraulic conductivity zones, Fig. 6.2. In Christensen and Cooley's study (1999) the hydraulic conductivity in zone no. 8 was held constant in the regression and in the succeeding prediction of state variables. For purposes of comparison the hydraulic conductivity in zone no. 8 was also held constant during this analysis. The hydraulic conductivity in the seven remaining zones was the target of calibration.

In the south-east part of the catchment there is a secondary aquifer, Fig. 6.2. The transmissivity was assumed to be constant within this aquifer

An aquitard was assumed to be present in regions where the estimated overload of till was larger than 1 m, see Fig. 6.3. The vertical conductivity was assumed to be constant within the aquitard.

The model setup consisted of two computational layers. The lower layer represented the main aquifer, while the upper layer represented both the secondary aquifer in the south-east part of the catchment and the aquitard where it was

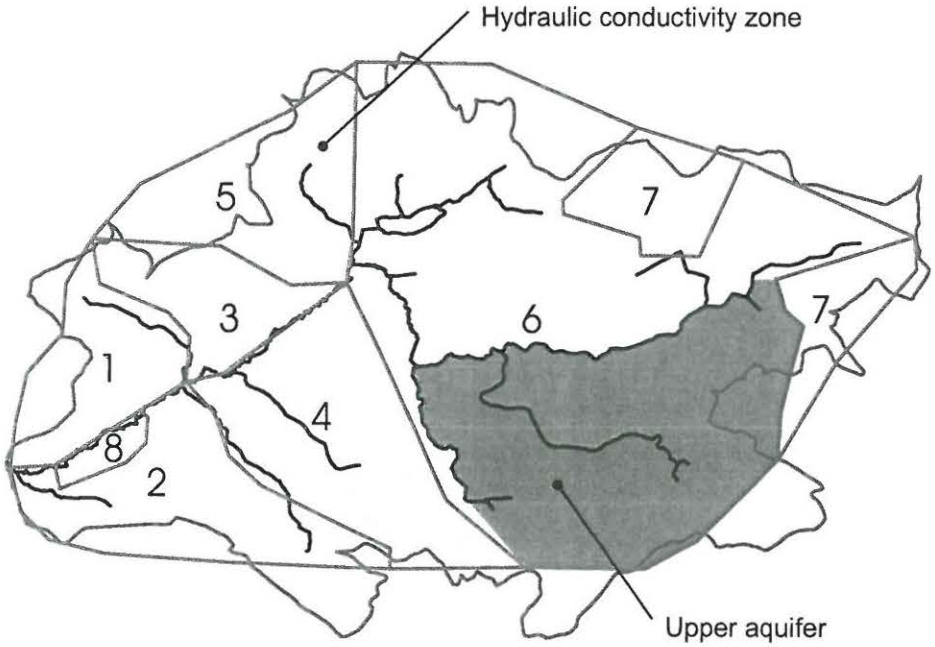


Figure 6.2: Hydraulic conductivity zones and location of the upper secondary aquifer. Numbers refer to the hydraulic conductivity zones

present. In the remaining areas the upper and lower layers were taken together.

In order to reduce stability problems in the calculation of heads and in the regression analysis, Christensen et al. (1998) considered all aquifers to be confined, even though this was not the case throughout the catchment.

6.1.2 Recharge

Three different types of recharge were calculated in three different ways respectively: (1) recharge directly to the lower model layer in regions with no aquitard, $RCH1$, (2) recharge to the upper secondary aquifer, $RCH2$ and (3) leakage from the aquitard to the lower aquifer. The heads in the aquitard were fixed to a level 3 m below the ground surface and the leakage was calculated on the basis of the head difference and the conductance between the upper and lower layers.

6.1.3 Streams

The streams presented in Figure 6.1 were included in the setup as general internal head boundary cells (McDonald and Harbaugh 1988). The interaction between the streams and the groundwater aquifers was governed by the hydraulic con-

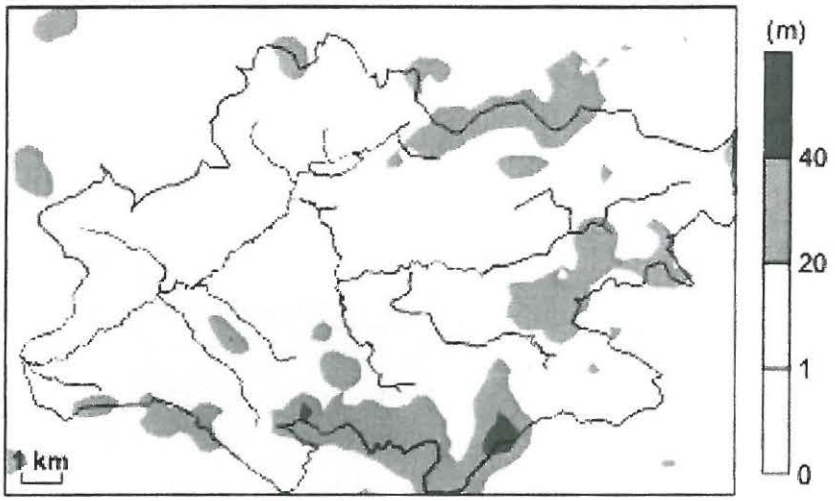


Figure 6.3: Thickness of till layer

ductivities of the stream bed and the difference between the water level of the stream and the groundwater potential. From the zonation of the hydraulic conductivities two stream bed hydraulic conductivities were suggested. These were constant within Christensen and Cooley's analysis (1999) as well as in this study.

6.2 Observation data

In the calibration procedure 64 hydraulic head observations and one stream flow observation were used. Because of their varying quality, the head observations were weighted unequally in the regression, on the basis of the inverse of the estimated prediction error. The head prediction error was estimated by Christensen and Cooley (1999) as ranging from 2 to 3.7 m and includes contributions from time effects, observation errors and errors due to small scale geological heterogeneity. The stream flow prediction error was estimated at 10 % of the observed value, Fig. 6.4. Wells nos. 1 to 56 are located in the main aquifer and wells nos. 57 to 64 are located in the secondary aquifer.

In the validation procedure 34 additional head observations (all in the main aquifer) and six additional stream flow observations were included, Fig. 6.5.

6.3 Monte Carlo simulations

Steady state simulations were performed using the groundwater model MODFLOWP (Hill 1992).

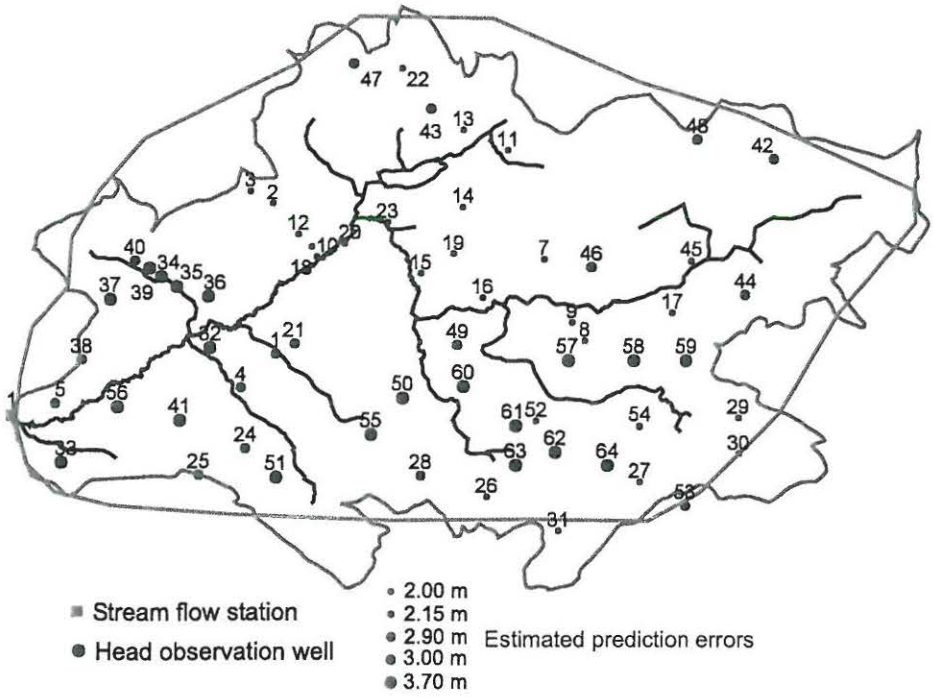


Figure 6.4: Observations used in calibration. Blue circles represent head observations and their diameter indicates the estimated prediction error ranging from 2 m to 3.7 m. Red squares represent stream flow gauging stations.

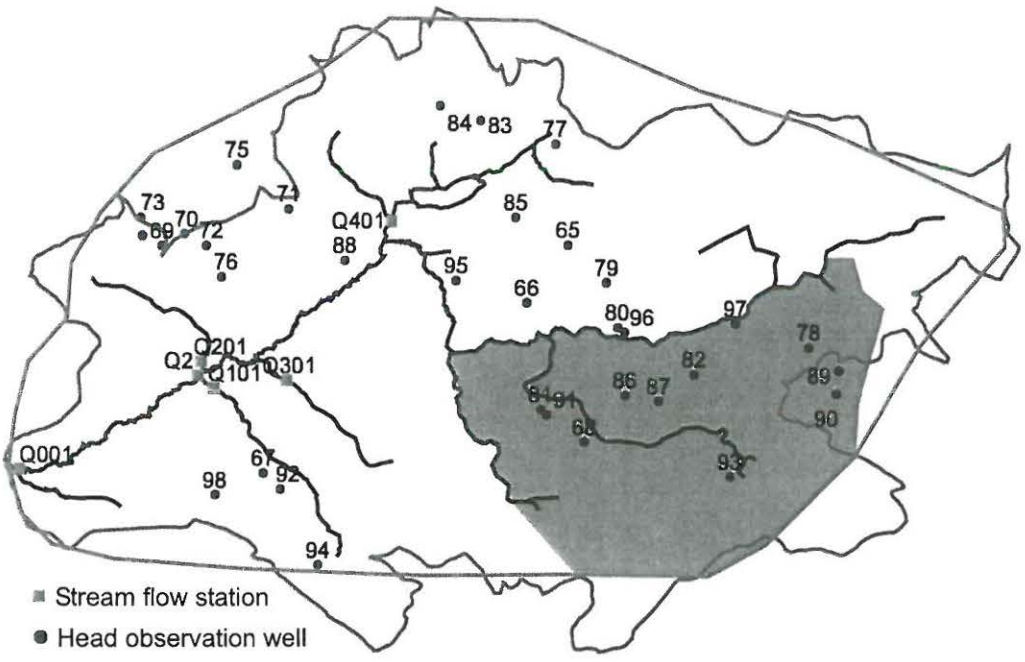


Figure 6.5: Observations used in validation. Blue circles represent head observations; red squares are stream flow gauging stations.

MODFLOWP was slightly modified by the author in order to perform Monte Carlo simulations. Due to the large number of simulations it was very essential that the Monte Carlo simulations were an integrated part of the MODFLOWP code; the initial part of each simulation was therefore skipped. The simulations were performed in a PC cluster using six standard PCs with CPUs ranging from Pentium II 400 MHz to Pentium III 933 MHz PCs. The average clock frequency was 700 MHz. A job control module was built into MODFLOWP, which made it possible to have an unlimited number of computers drawing from the same pool of parameter realisations. The PC cluster was able to perform 1.2 million simulations per day.

In total, seven horizontal hydraulic conductivities, $K1 - K7$ (zone nos. 1-7), one vertical hydraulic conductivity, Kv (aquitard), one transmissivity in the secondary aquifer, TU and two recharge values, $RCH1$, $RCH2$ were the target of calibration.

Sample intervals for the eleven parameters were constructed and 30 million random realisations or sets of the 11 parameters were carried out.

Initially very wide uniform sampling intervals (The A intervals on Fig. 6.6) were set, and on this basis 10 million simulations were performed. The GLUE analysis based on these simulations indicated that it would be advantageous to reduce the sampling intervals, and the shorter B intervals were therefore introduced. The reduction of the parameter intervals from A to B reduced the size of the sampling space by a factor of 1500 and thereby increased the number of simulations in the accepted region by a similar factor.

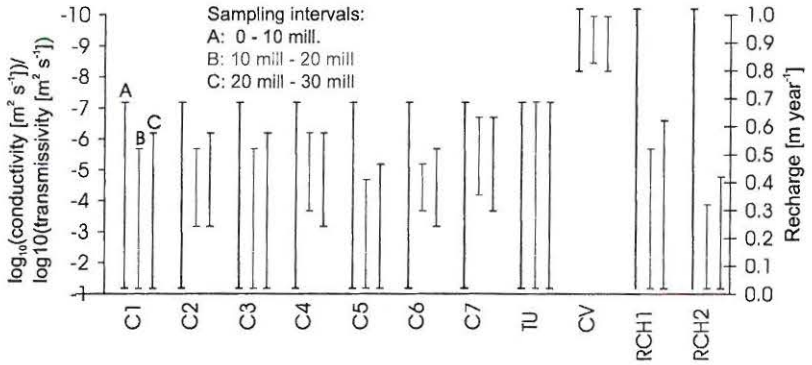


Figure 6.6: Sample space for the 11 parameters targeted for calibration

On the basis of the B intervals another 10 million simulations were carried out and a GLUE analysis was performed on all 20 million simulations. This indicated that for certain parameters a further adjustment of the intervals was needed. The C intervals were therefore introduced and on this basis the final 10 million simulations were performed, Fig. 6.6.

6.4 GLUE analysis

In this study the observation point likelihood for all observation points was calculated on the basis of the Gaussian likelihood functions (see Chapter 4):

$$L\left(\boldsymbol{\theta}|\psi_i^*\right) = \frac{1}{\sqrt{2\pi}\sigma_{\psi_i^*}} e^{-\left(\frac{\left(\psi_i^* - \psi_i(\boldsymbol{\theta})\right)^2}{2\sigma_{\psi_i^*}^2}\right)} \tag{6.1}$$

The rejection level was set at $3\sigma_{\psi_i^*}$. The Gaussian likelihood function requires an estimate of expected head and stream flow errors. Initially the estimates from Christensen et al. (1998) were used; see also Fig. 6.4. The estimated head error varied from 2.0 - 3.7 m and the estimated stream flow error was set at 10 % of the observed value.

The point likelihood measures were aggregated into a global likelihood measure using the geometric mean inference function:

$$L\left(\boldsymbol{\theta}|\boldsymbol{\psi}^*\right) = \sqrt[N_{obs}]{\prod_{i=1}^{N_{obs}} L\left(\boldsymbol{\theta}|\psi_i^*\right)} \tag{6.2}$$

Here N_{obs} is 65 (one stream flow observation and 64 head observations).

Using the expected head and stream flow observation errors from Christensen et al. (1998) and the point likelihood function, rejection level and inference function presented above, the GLUE analysis resulted in 404 accepted simulations - each with an individual likelihood.

The accepted simulations were used to calculate the discrete likelihood distribution curve for all observation points. From the likelihood distribution curves the expected value, the upper and lower 95% prediction intervals and the minimum and maximum accepted values were found.

6.4.1 Head results

The head results, together with observed values and the most likely simulated value, are presented in Figs. 6.7 and 6.8.

Fig. 6.7 presents the results of the 64 head observations used in the calibration procedure. Of these 64 observations sixteen (25%) fell outside the 95% prediction intervals and nine (14%) fell outside the minimum/maximum interval.

Fig. 6.8 presents the results from the 34 head observations used in the validation procedure.

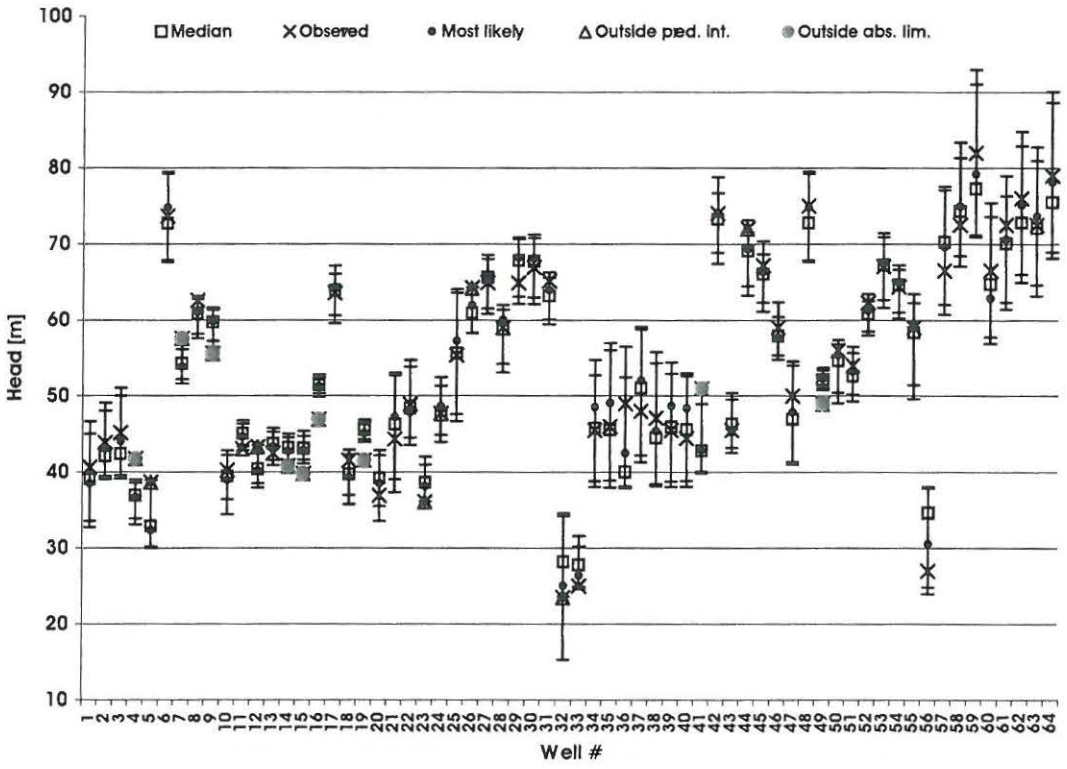


Figure 6.7: Observed and simulated heads used in calibration. Inner error bars are 95% prediction intervals and outer error bars are minimum and maximum head in accepted simulations. Observations that fall outside the 95% prediction interval are marked in blue and observations that fall outside the minimum/maximum limits are marked by red.

Of these 34 head observations, ten (29%) fell outside the 95% prediction intervals and nine (24%) fell outside the minimum/maximum interval.

A small reduction in predictive capability was noted in the case of the additional observations that were not used in the calibration.

It is interesting to note that a considerable number of observations fell outside the 95% prediction interval, while at the same time we had very wide (up to 20 metres) 95% prediction intervals for some observations. In general the prediction intervals varied from 2 to 20 metres.

The results from Figs. 6.7 and 6.8 are visualised in Fig. 6.9. Observations outside the 95% prediction intervals are shown with varying symbol size, depending on how far outside the 95% prediction intervals the single observation was located.

The observation points with poor prediction capacity tended to be grouped together. In the centre of the catchment heads were in general overestimated. These results indicate certain structural problems in the conceptual model.

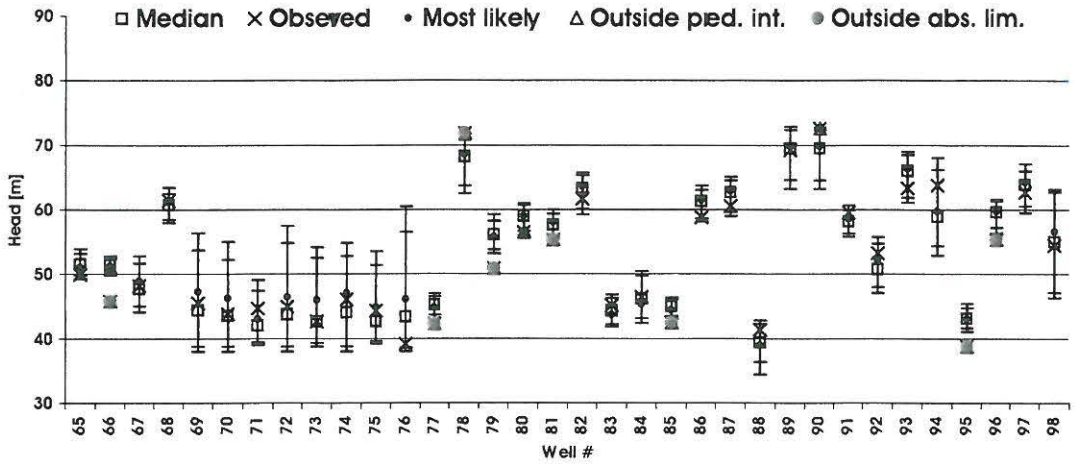


Figure 6.8: Observed and simulated heads used in validation. Inner error bars represent 95% prediction intervals and outer error bars represent minimum and maximum head in accepted simulations. Observations that fall outside the 95% prediction interval are shown in blue and observations that fall outside the minimum/maximum limits are shown in red.

It was concluded from this that the present GLUE analysis does not result in a satisfactory validation of the head predictive capability.

Christensen and Cooley (1999) performed a validation of the prediction of heads in the additional 34 head observation points. They found only one head observation outside the non-linear prediction interval. The calculation of linear prediction intervals is presented in Section 3.9. The calculation of the state variable statistics depends on the estimate of the true observation error (the last term in Eq. 3.22). The calculation of non-linear prediction intervals is based on similar conditions.

Thus the successful validation in Christensen and Cooley (1999) was closely connected to the distribution of the estimated true observation errors. Christensen and Cooley (1999) distributed the estimate of the true observation errors proportionally in relation to the estimated observation error of heads used in the calibration. Regions where the calibration head observations are associated with large estimated errors will have correspondingly large errors in the head prediction.

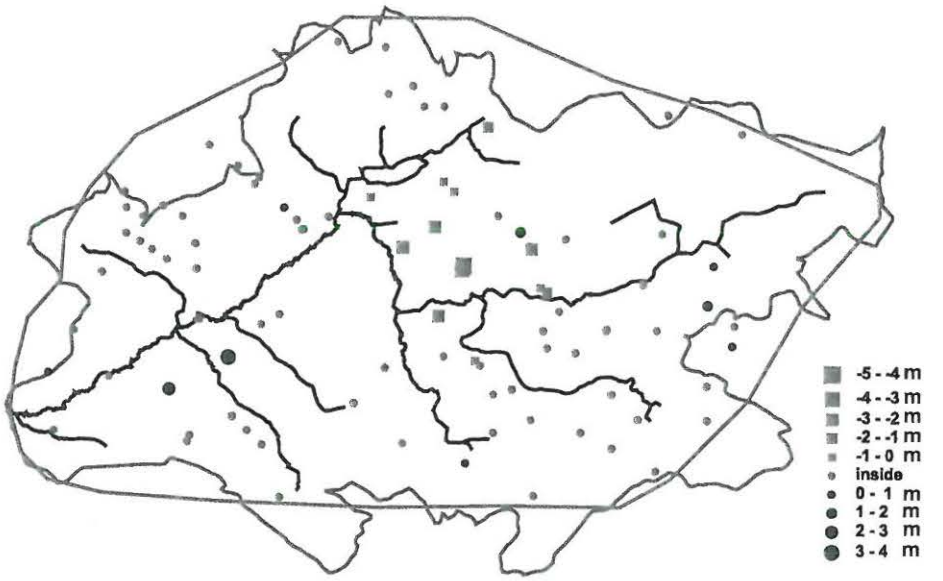


Figure 6.9: Predictive capability of simulating heads (calibration and validation data set). Red squares and blue circles indicate wells where the observation is below the lower 95% prediction limit and above the upper 95% prediction, respectively. The symbol size indicates how far the observed value is from the 95% prediction limits. Green points indicate wells where the observed value is inside the 95% prediction limits.

6.4.2 Stream discharge results

The stream discharge results, together with observed values and the most likely simulated value, are presented in Fig. 6.10.

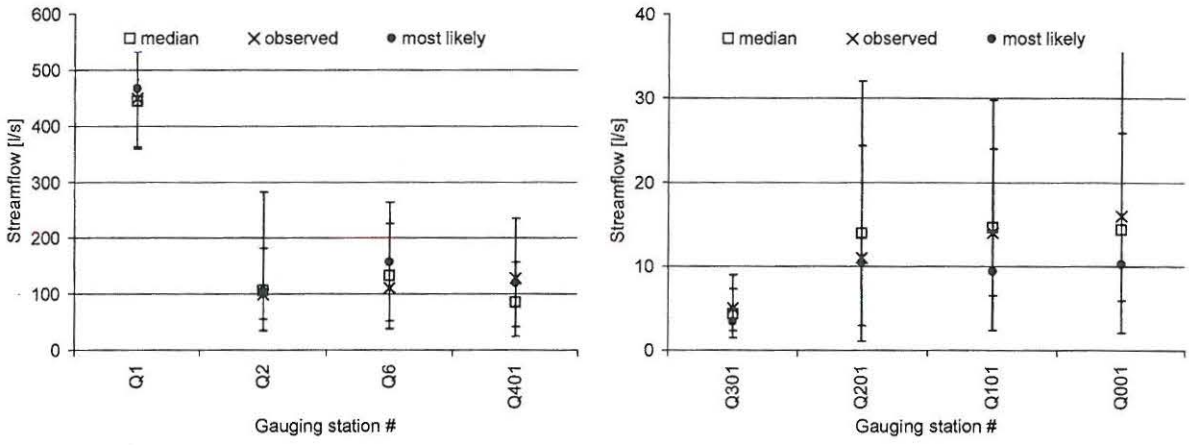


Figure 6.10: Observed and simulated stream flows. Inner error bars represent 95% prediction intervals and outer error bars represent minimum and maximum head in accepted simulations. Station Q1 was used in the calibration and the remaining stations belongs to the validation data set

The prediction intervals for all stream discharges cover the observed values, which confirms that the predictive capability depends on the scale of the process under consideration. Stream flow is in general more predictable than heads, Table 1.1.

6.4.3 Parameter estimates

From the GLUE analysis 404 simulations were accepted. The parameter likelihood scatter plot is presented in Fig. 6.11.

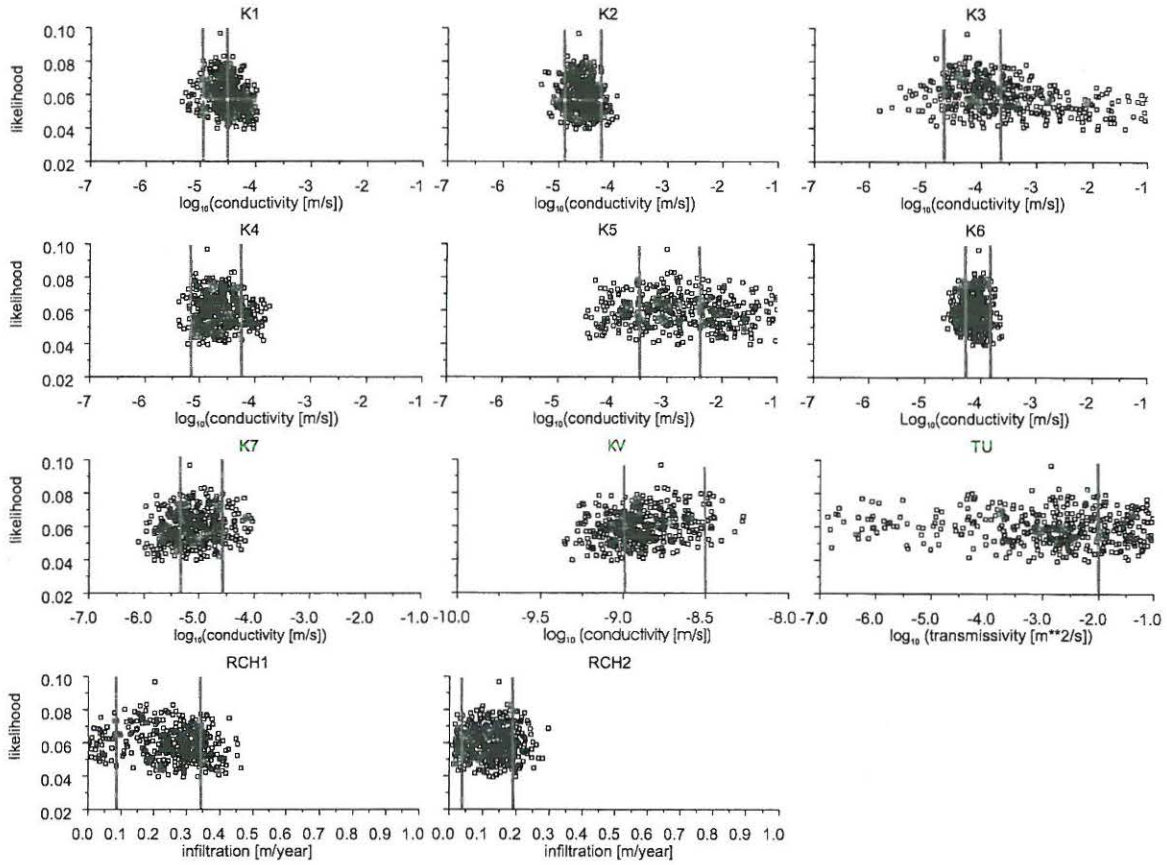


Figure 6.11: Parameter likelihood scatter plot - Gaussian likelihood function and geometric mean inference function. Red lines indicate the upper and lower bounds of the 95% prediction intervals found by Christensen and Cooley (1999)

The likelihood value corresponding to the optimal parameter set found by Christensen and Cooley (1999) is included in the scatter plot. This parameter set yields the highest likelihood values, and from this there is quite a large jump to the second highest. This indicates that the likelihood surface is steep in the region around the optimal parameter set found by regression. From this we can conclude that the sample density is too small to resolve the surface in this region.

The non-linear confidence intervals from the regression presented in Christensen and Cooley (1999) are indicated by the red lines in Fig. 6.11. The GLUE-based intervals of accepted simulations are larger for all parameters than the intervals from Christensen and Cooley (1999). However, This was to be expected given that the confidence intervals reflect uncertainties only in the parameters and not in the model.

6.4.4 Reducing the GLUE rejection criteria

This section examines whether the poor predictive capability of GLUE revealed in this case was the result of an over-strict rejection criteria. The expected error on head and stream flow was increased by 50%, the GLUE analysis was then repeated and 100,500 simulations were accepted. The head prediction intervals are presented in Figs. 6.12 and 6.13.

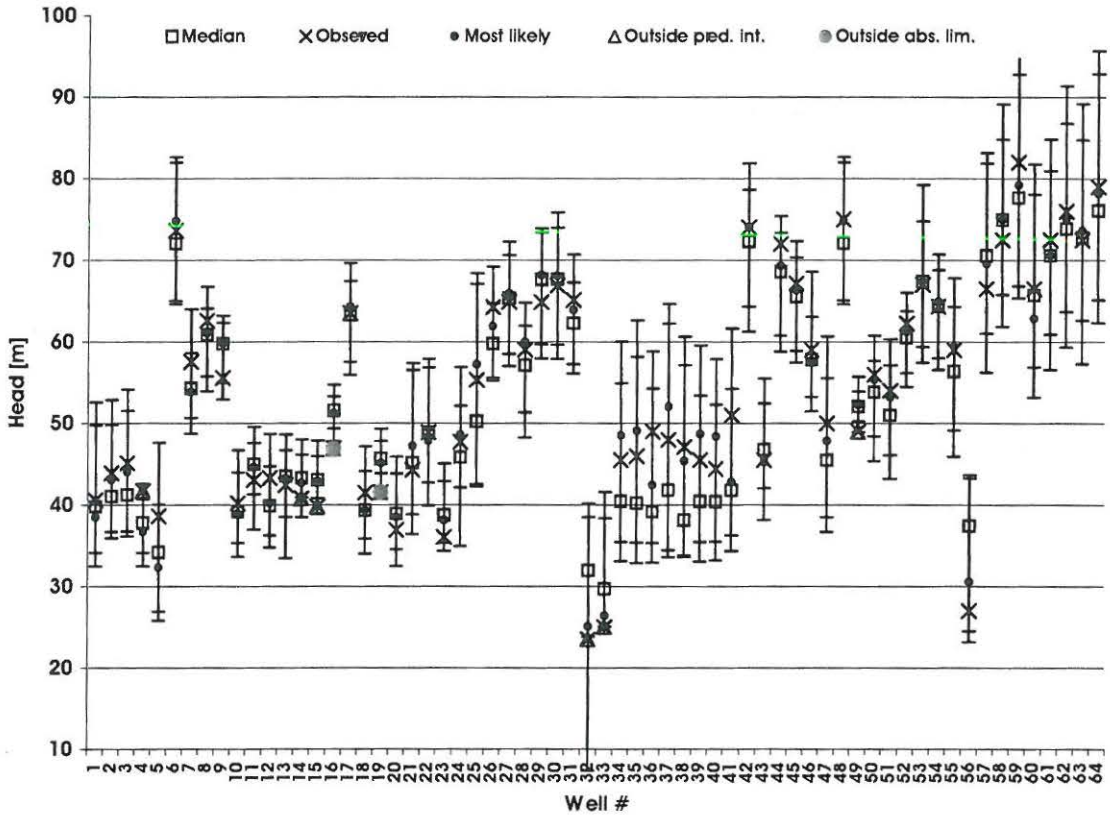


Figure 6.12: Observed and simulated heads used in calibration. Inner error bars prepresent 95% prediction intervals and outer error bars represent minimum and maximum head in accepted simulations. Observations that fall outside the 95% prediction interval are shown in blue and observations that fall outside the minimum/maximum limits are shown in red

Of the 64 observations used in the calibration, eight (13%) fell outside the 95%

prediction interval and two (3%) fell outside the minimum/maximum interval. Of the additional 34 head observations, four (12%) fell outside the 95% prediction

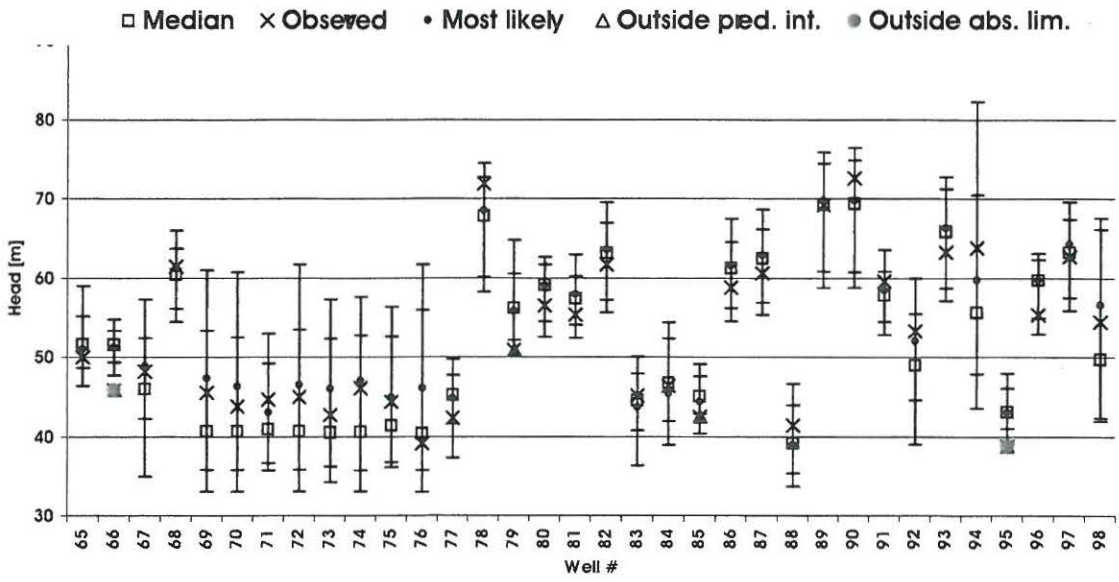


Figure 6.13: Observed and simulated heads used in validation. Inner error bars represent 95% prediction intervals and outer error bars represent minimum and maximum head in accepted simulations. Observations that fall outside the 95% prediction interval shown in blue and observations that fall outside the minimum/maximum limits are shown in red

interval and two (6%) fell outside the minimum/maximum interval.

With an even weaker rejection criterion 12% of the observations fell outside the prediction interval, while at the same some of the prediction intervals were very wide (up to 28 metres), which makes the estimate less useful.

6.4.5 Convergence

Finally, we examined whether the poor predictive capability was a result of non-convergence in the estimate.

In Fig. 6.14 the development of the expected head, the 95% prediction interval and the minimum/maximum intervals are presented as a function of the number of accepted simulations for six representative wells.

If we examine this figure the solution seems to stabilise within 50 - 100 accepted simulations. If we assume the 95% prediction interval to be close to the true interval with 400 accepted simulations, it is quite simple to calculate the necessary number of accepted simulations, $N_{e^{95\%}}$, in order to obtain a certain precision, $e^{95\%}$. We start from the estimated value of the upper and lower 95% predic-

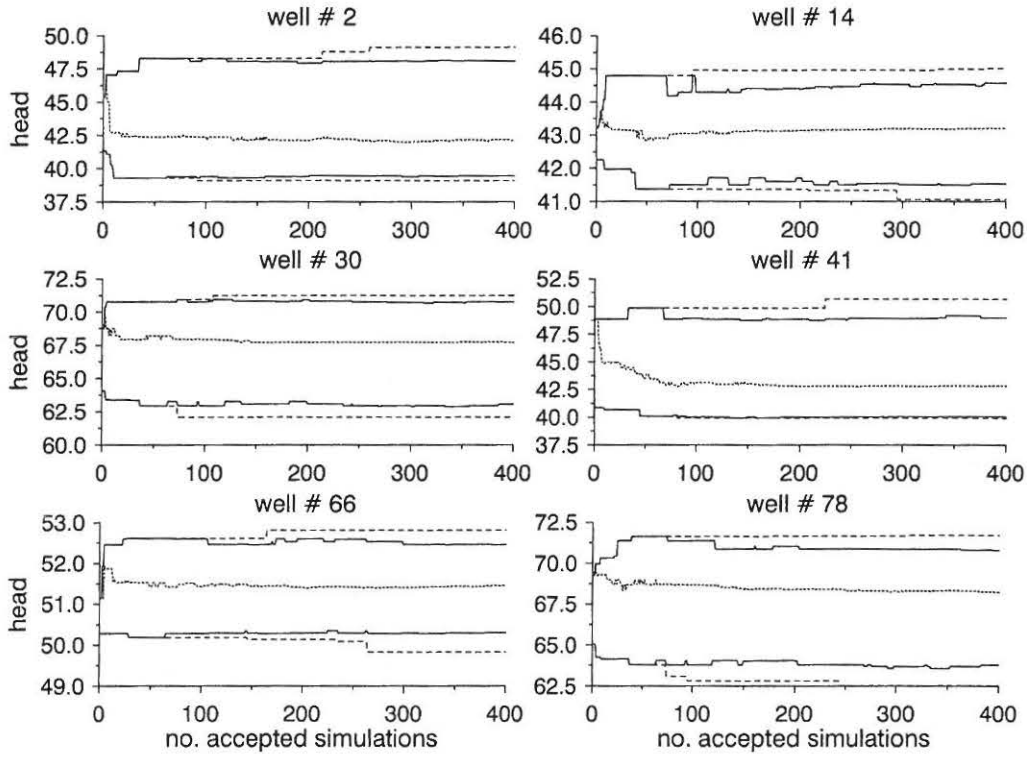


Figure 6.14: Expected head (dotted line), 95% prediction interval (continuous lines) and minimum/maximum interval (dashed lines) as a function of the number of accepted simulations

tion interval at 400 accepted simulations and reduce the number of accepted simulations, $N_{e^{95\%}}$ until

$$e^{95\%} \leq \max\left(\left|h^{2.5\%}(N_{e^{95\%}}) - h^{2.5\%}(N_{acc})\right|, \left|h^{97.5\%}(N_{e^{95\%}}) - h^{97.5\%}(N_{acc})\right|\right) \tag{6.3}$$

The result for well no. 30 with $e^{95\%}$=0.25m is illustrated in Fig. 6.15. Close to 200 accepted simulations ($N_{0.25\mathrm{m}}$) are needed in order to achieve an estimate with precision better then 0.25 m.

If we consider all observation wells the average number of necessary accepted simulations can be found as

$$\overline{N}_{e^{95\%}} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} N_{e^{95\%},i} \tag{6.4}$$

The necessary number of accepted simulations in order to obtain a maximum error of 0.1 m, 0.25 m and 0.5 m of the 95% prediction interval is found to be 308, 110 and 40 respectively, based on the six wells presented in Fig. 6.14.

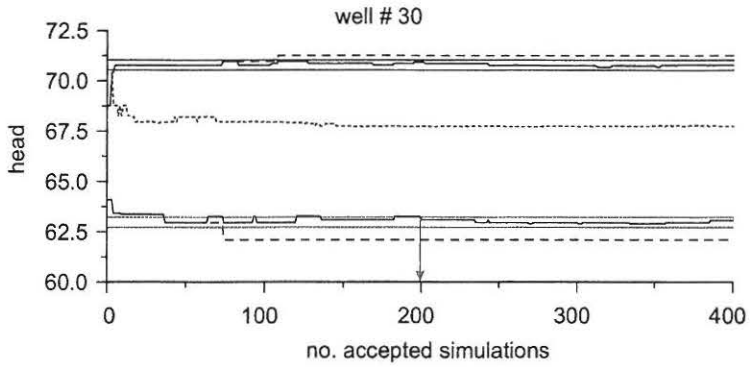


Figure 6.15: Example of the determination of the necessary number of accepted simulations. Red lines represent the 95% lower prediction interval ± 0.25 m and blue lines represent the 95% upper prediction interval ± 0.25 m. The necessary number of accepted simulations is defined as the first point at which these lines cross when sweeping from right to left.

It should be noted that the result from the convergence analysis will depend on the likelihood and inference functions. A more restrictive (steeper) likelihood allocation will probably decrease the convergence speed.

On the basis of the likelihood and inference functions applied, it can be concluded that convergence problems do not explain the poor predictive capability of the GLUE analysis.

It has been shown that the prediction of the uncertainty of state variables converges to a reasonable level with a relatively low number of accepted simulations. However, the reader should note that the estimation of the parameter response surface requires a much larger number of accepted simulations.

6.5 Summary and conclusion

The GLUE methodology was applied to a regional aquifer system, which had earlier been the target for calibration and uncertainty predictions within a regression framework (Christensen et al. 1998; Christensen and Cooley 1999). The present analysis is based on exactly the same conceptual model and numerical formulation of the flow problem. The main purpose of the study was to predict uncertainties in groundwater heads and stream flows and to compare the results archived with those obtained from the regression analysis.

In the calibration procedure 64 head observations and one stream flow observation were used. In Christensen and Cooley (1999) the expected observation errors were evaluated for all observations. The head observation error was esti-

mated to range from 2.0 to 3.7 metres and the stream flow observation error was estimated at 10% of the observed value.

The GLUE analysis was applied to the setup in three steps, each including 10 million Monte Carlo simulations. Initially very wide parameter intervals were sampled, but on the basis of the GLUE analysis these parameter intervals were subsequently reduced.

The GLUE analysis was based on the 30 million simulations performed and for each simulation a point likelihood measure was calculated, using the Gaussian likelihood function, for the 64 head observation points and the one stream flow observation station. The rejection level was set at three times the estimated observation error. The global likelihood measure was calculated from the 65 point likelihood measures using the geometric mean inference function.

The GLUE analysis resulted in 404 accepted simulations, each with an individual likelihood. On the basis of the accepted simulations the likelihood distribution curves were calculated for the 65 observations used in the GLUE analysis and for 34 additional head observations and seven additional stream flow observations. From the distribution curves the 95% prediction intervals and the minimum and maximum value were calculated.

The results showed that over 26% of the observations were outside the 95% prediction interval, and over 18 % were outside the minimum and maximum limits. It was concluded that the GLUE analysis did not result in a satisfactory validation result.

It was found that the observation points where the validation failed were grouped together. In the centre of the catchment heads were in general overestimated. The result shows very clearly that the observation errors are interdependent, and it was suggested that the reason for the poor validation results might be found in conceptual model errors.

A subsequent investigation was made to see whether the poor validation was the result of an over-strict rejection criterion. The expected observation errors were increased by 50% and another GLUE analysis was performed that resulted in 100,500 accepted simulations. It was found that 12% of the head observations were still outside the 95% prediction intervals and 4% were outside the minimum and maximum limits. At the same time the results showed very wide prediction intervals of up to 28 metres.

From this it may be concluded that the GLUE analysis was unable to predict uncertainties in the simulated heads in the given model setup and that the conceptual model has to be reconsidered in order give satisfactory results.

CHAPTER 7

Summary and conclusion

Groundwater models play an important role in the assessment of groundwater resources and in the protection of these. Considerable costs are associated with the decisions taken on the basis of groundwater model simulations, and it is therefore essential that groundwater models are "reliable".

However, reliability depends to a great extent on the data basis available for model construction and calibration. In most situations this data basis is very incomplete, and certain hydrologists therefore consider the models based upon them to be "unreliable", which - from a deterministic point of view - is often true. These criticisms need to be met, and since it is difficult to increase the reliability of the models by expanding the data basis, we must try instead to quantify the uncertainty inherent in the outcome of groundwater models. In general, uncertainty is undesirable, but not knowing the size of it is even more so.

7.1 Uncertainty estimation

The author has identified three approaches to uncertainty modelling: 1) forward stochastic approach, 2) conditioning stochastic approach and 3) regression approach.

The forward stochastic approach refers to the ordinary stochastic methods in which uncertainty in the system input is transmitted through the model to uncertainty in the system state variables. The solution is not conditioned on any observed system state variables.

The conditioning stochastic approach refers to those models in which observed system state variables are included and the uncertainty estimated is conditioned on the observations and the prior statistics of the input.

The regression approach are designed to find a parameter set that optimises a certain model fit criterion. The regression analysis provides approximate statistics on the estimated parameters and on the system state variables.

In general it is accepted that it is necessary to calibrate the models in order to reduce uncertainty in model outcome. It is however very rare that the data basis is of a sufficiently high quality that a forward stochastic approach can be applied directly. This thesis therefore concentrates on the conditioning stochastic approach and the regression approach.

In order to evaluate the advantages and disadvantages of the various methods for estimating uncertainty, it is very important to understand the possible sources of uncertainty. Section 3.2 contains a description of the different types of observation data that can be used in the calibration, and the different sources of mismatch between observed and simulated state variables.

The errors that give rise to mismatch between observed and simulated state variables may be devided into two types: observation errors and other errors. Observation errors consist of all the contributions that are independent of the model and relate only to observation. E.g. head observation errors are the errors associated with the manual reading and registration of water levels, barometric fluctuations, the determination of the well reference level, etc. Other errors must be associated with the model setup and with the numerical formulation of the specific flow problem, and as such may be classified as model errors. Observation errors are typically much smaller than model errors, and it is important to be aware of this, since the two types of error influence model outcome in different ways. It is characteristic of observation errors as defined above that their impact on model outcome diminishes as the number of observations increases, while the influence of model errors on model outcome is in general invariant in relation to the number of observations.

It should be noted that the definition of observation errors is often more wide-ranging than that given above, and may include all sources of error that over large areas "act" as observation errors (non-biased and Gaussian distributed errors).

Prior to any uncertainty estimation it is important to recognise which processes the groundwater modelling procedure includes and, even more important, which processes it does *not* include. The processes that are not included constitute the source of all conceptual model errors. By recognising the error contributions of the processes that are not included, the hydrologist will be able to formulate goals

for the precision of the model outcome. If these goals are not achieved, either the conceptual model itself, or the identification of the processes that were excluded, and which may therefore have contributed to error, must be reconsidered.

7.2 The classical regression approach

The classical regression approach as presented in Chapter 3 provides confidence intervals on estimated parameters and in principle on all system state variables. Confidence intervals represent the uncertainty that originates from observation errors and thus constitutes only a fraction of the total uncertainty.

In classical regression theory the total uncertainty is given in terms of prediction intervals. To the best of the author's knowledge prediction intervals have been formulated only for state variables such as heads and river discharges, on the assumption that the true error associated with the prediction is known. Even where this is the case, however, there is no obvious relationship between the uncertainty of head/stream flow on the one hand, and that of groundwater pore velocities on the other. Frequently, it is in relation to groundwater pore velocities that uncertainty estimates are wanted.

Another restriction within the classical regression approach is that the estimation has to be well-posed. Well-posedness is characterised by an identifiable, unique and stable solution, and in principle can always be obtained by reducing the number of unknown models, parameters and variables. This will however affect the amount of model errors, as discussed in section 3.3.

7.3 The GLUE methodology

The GLUE methodology presented in Chapter 4 can be characterised as a *conditioning stochastic approach.* The fundamental idea behind the methodology is that a wide range of models, parameters and variables are likely to be simulators of the system under consideration and should therefore be considered prior to modelling. A large set of random realisations of these models, parameters and variables are produced and Monte Carlo simulations are performed. The GLUE methodology aims to categorise the space of models, parameters and variables into a subspace of behavioural/accepted solutions and a subspace of non-behavioural/rejected solutions. The space of models, parameters and variables is divided on the basis of more or less subjective likelihood measures, and different degrees of membership are given to the behavioural solutions - again based on subjective likelihood measures.

The uncertainty of model outcome is represented by the likelihood surface in the space of behavioural solutions. Each accepted set of models, parameters and variables contributes to a realisation of model outcome, and from all accepted sets of models, variables and parameters density functions can be found.

The "uncertainty" formulated in the GLUE methodology is not uncertainty as defined in the classical regression framework, but a more subjective impression relating to the individual hydrologist.

The key issue in the GLUE methodology is the calculation of likelihood measures. In the present study a Fuzzy logic approach is taken, involving an evaluation of the likelihood in each observation point, which is referred to as the *point likelihood measure*. For all observation points the point likelihood measure is aggregated into a global likelihood measure by using an *inference function*.

A number of global likelihood measures are presented and it is argued that the point likelihood measure should be closely connected to the expected error in the observation, and that the inference function should be of the *and*-type - i.e. that all point likelihood measures must be positive in order to produce a positive global likelihood measure.

7.3.1 Synthetic case study

Chapter 5 presents capture zone modelling in a synthetic stationary integrated 3-dimensional flow problem including river, surface and groundwater flow. A synthetic setup was chosen in order to be able to evaluate the model outcome against a known solution.

A reference model was formulated with three geological layers - a lower aquifer, an aquitard with a sandy window and an upper aquifer. The conductivities in the geological layers were heterogeneous and were modelled as random fields.

Two geological models were formulated within the conceptual model. Both models consisted of three geological layers with homogeneous hydraulic conductivities. Model A contained an aquitard in the entire catchment, while model B contained a sandy window in the aquitard.

The reference model contained geological heterogeneities that were not described in the conceptual model. This corresponds to the situation in most groundwater model applications, where lack of data forces the hydrologist to make simplifications.

The reference model and the conceptual model were identical in all other respects, and thus the geological heterogeneities were the only possible source of

mismatch between the reference model and the conceptual model. The expected error in the head observation points was estimated on the basis of on the degree of heterogeneity, and the expected error in the river flow observation was estimated on the basis of general assumptions.

The conceptual model was adapted to the FLEXFLOW model presented in Chapter 2. The construction of FLEXFLOW was motivated by the lack of flexibility in existing finite difference models with regard to numerical discretisation, representation of geological units, river integration and surface, river and groundwater interaction - all of which are important in relation to reducing model errors. Furthermore it is essential in the GLUE methodology to optimise computational speed. In FLEXFLOW the flow processes are fully integrated, and the Monte Carlo simulation is likewise an integrated part of the model and both aspects help to increase computational speed.

In the GLUE analysis seven parameters were considered unknown and were the target for estimation: the two geological models, the precipitation and five hydraulic conductivities. In total 200,000 realisations of the seven unknown parameters were established as a basis for FLEXFLOW simulations.

A number of GLUE analyses were performed, each resulting in a number of accepted simulation/parameter sets, and on the basis of the accepted simulations capture zone likelihood distribution maps were produced. These maps were then compared with the reference capture zone.

The main conclusions were:

- Both head and river discharge observations needed to be included in the GLUE analysis of the present setup in order to obtain a capture zone distribution that was similar in shape and extension to the reference capture zone.
- Likelihood measures based on the estimated observation errors result in a 95% prediction zone that covers the reference capture zone.
- The capture zone distribution seems invariant in relation to the type of applied point likelihood function.
- The capture zone was highly dependent on the inference function. The *and-type* inference functions (geometric mean and product inference) seem to have considerably greater predictive capability than the max. and arithmetic mean inference functions.
- A 25% increase in the rejection level has no significant impact on the shape of the capture zone distribution, whereas a 100% increase of the rejection level has a significant impact on the capture zone distribution.

- A validation was performed in twelve head observation points and two river discharge stations, showing a decrease in the predictive capability in a region close to the abstraction well. Moreover the likelihood distribution of the river discharge showed weak predictive capability. Three observations were outside the band of simulated values. This indicated that the likelihood measure was too restrictive, and for this reason the rejection level was increased by 25%. As a result, all observations were placed well within the simulated likelihood intervals.

- The GLUE analyses that were considered most reliable (e.g. analysis 1.4) all resulted in prediction of the wrong geological model. For this specific case study, we may conclude that geological heterogeneities may be a disturbing factor in the identification of the correct geological model.

7.3.2 Regional aquifer case study

Chapter 6 presents the application of the GLUE methodology applied to a regional aquifer system which had previously been target of parameter and uncertainty estimation within the classical regression framework (Christensen et al. 1998; Christensen and Cooley 1999). The present work was conducted on exactly the same conceptual and numerical model as that presented in Christensen et al. (1998).

The purpose of the GLUE analysis was to predict uncertainties in groundwater heads and stream flows and to compare the results obtained with those obtained from the regression analysis.

A GLUE analysis was performed on 64 head observations and one stream flow observation. A Gaussian point likelihood measure, with estimated observation errors ranging from 2 to 3.7 m on heads and 10% of the observed stream discharge, was calculated in the 65 observation points, and the point likelihood measures were aggregated into a global likelihood measure using the geometric mean inference function.

On the basis of the accepted simulations likelihood distribution curves were calculated for all the 65 observations used in the analysis, and for an additional 34 head and seven stream discharge observations.

The stream discharge results were successfully validated for all stations, while 25% of all head observations were not contained in the 95% prediction interval. An investigation was made to see whether the poor validation of heads was a result of an over-strict likelihood measure. For this reason the expected observation errors were increased by 50% and a new GLUE analysis was performed. It was found that 12% of all head observations still remained outside the 95%

prediction interval, while at the same time certain prediction intervals became extremely wide, going up to 28 m.

Fig. 6.9 presents the distribution of the observation points where validation failed, and it can be seen that the points where validation failed were grouped together. It was concluded that there were strong indications of conceptual model errors in the setup.

With the current setup it was impossible using the GLUE methodology to quantify the predictive uncertainty. This conclusion contrasts with the very successful validation presented in Christensen and Cooley (1999), which is based on classical regression theory.

The reasons for these contrasting conclusions are to be found in the nature of GLUE. The likelihood measures are such that a certain amount of uncertainty in the simulated state variables is allowed, and combinations of models, parameters and variables that fulfil the restrictions within the likelihood measure are accepted. In this way GLUE projects all uncertainty onto the unknown models, parameters and variables in the given problem. In the present case study the conceptual model errors are presumably of such a size that the uncertainty cannot be accounted for in terms of the parameters and variables that are defined as unknown.

The GLUE methodology did however provide very useful information on possible weaknesses in the model setup, and the present analysis is expected to very useful in the future development of the Gjern model.

7.4 Remarks on the GLUE methodology

The application of the GLUE methodology in the present case studies has shown that the approach is feasible for integrated stationary models on the field scale, and stationary groundwater models on the catchment scale with up to eleven unknown parameters.

When applying GLUE to larger models and more parameters the computational burden is assumed to be linear, increasing by the number of computational nodes in the model, and non-linear, increasing by the number of additional parameters.

The GLUE methodology is computationally intensive and this is one of the major hurdles to overcome when considering a wider application of the methodology. The future application of the methodology may very well depend on the development of more efficient sampling methods (e.g. the exclusion of certain regions in the parameter space on the basis of a preliminary search). More efficient

sampling methods, however, may very well constitute a threat to the freedom from prejudice that is essential to the GLUE methodology, since it is based on the premise that a wide range of combinations of models, parameters and variables must, until otherwise proven, be regarded as likely simulators of the system under consideration. In the author's opinion, however, it may be necessary to sacrifice some freedom in this regard in order to make sampling more efficient.

Beven and Binley (1992) acknowledge the computational burden involved in the GLUE methodology, but put faith on future increases in computational power. Since the work of Beven and Binley was published, the clock-frequency of a standard PC has increased at least 50 times. The problem is, however, that the increase in computational speed in general has not been used to improve uncertainty estimation but instead to build more and more complex models. In view of this, one might ask whether it is time now to freeze the development of model complexity and instead use the gain in computational efficiency to improve reliability of our estimations.

Many groundwater problems require the application of transient models. The GLUE methodology has so far mainly been applied to stationary groundwater models, and there is a need therefore to gain more experience in applying the method to transient groundwater problems. I believe however that the move to transient models should be taken step by step, and within this framework GLUE analyses on a stationary approximation of the transient problem may be a very important ingredient.

APPENDIX A

Reliability methods

This Appendix gives an introduction to the field of reliability estimation, describing in particular Monte Carlo simulation, Latin Hypercube simulation and the First Order Reliability Method. The purpose of all these is to transmit the uncertainty in the model input parameters through a deterministic ground water model to the desired output parameters, which in the case of a groundwater model are potential heads, solute concentrations, stream flows, capture zone areas, etc. It is assumed that the statistics of the input are known and that the numerical groundwater model is error free.

Monte Carlo (MC) and Latin Hypercube (LH) simulation are among the so-called simulation methods, where a large number of model simulations are performed and discrete probability density functions are obtained for output variables of interest. The First Order Reliability Methods (FORM) are designed for estimating the probability of a given event, e.g. groundwater nitrate concentration exceeding a given limit value.

A.1 Reliability

Let us consider the most simple decision-making problem. The parameter s describes the environmental load (concentration, water level, etc) and the parameter r describes the environmental resistance/capacity (limit value, critical water level, etc.) A critical situation occurs if the load exceeds the resistance/capacity. For a given realisation of the load parameter s and the resistance parameter r the situation is easily evaluated:

$$\begin{array}{l} r - s \leq 0 \quad \text{critical situation (failure)} \\ r - s > 0 \quad \text{non} - \text{critical situation (safe)} \end{array} \tag{A.1}$$

We define the failure function g as dividing realisations of r and s into a failure set w_f and a safe set w_s, the two sets being separated by the failure surface or limit state surface, which is defined as

$$g\,(r, s) = r - s = 0 \tag{A.2}$$

or in more general terms as

$$g\,(\boldsymbol{\theta}) = g\,(\theta_1, ..., \theta_n) = 0 \tag{A.3}$$

where $\boldsymbol{\theta}$ is a vector of parameters involved in the problem under consideration. Positive values of g correspond to safe states and zero or negative values of g correspond to failure states.

$g(\boldsymbol{\theta}) \leq 0$ in the failure region
$g(\boldsymbol{\theta}) > 0$ in the safe region

The failure function as it relates to the simple load/resistance problem described above is shown in Fig. A.1.

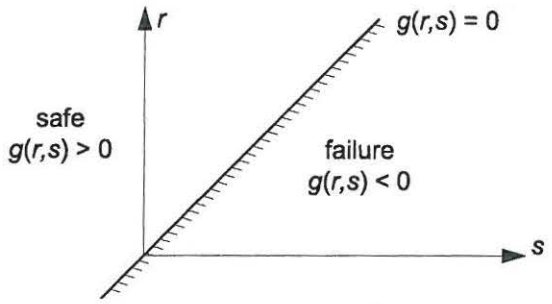


Figure A.1: Failure function

If the load parameter s is kept deterministic and the environmental resistance R is considered stochastic with a density function as shown in Fig. A.2, the situation becomes slightly more complicated.

We introduce the failure probability P_f to denote the probability that a critical situation (failure mode) will occur. The probability of failure is given as

$$P_f = P(\mathsf{R} - s \leq 0) = P(g\,(\mathsf{R}, s) \leq 0) = \int_{-\infty}^{s} f_{\mathsf{R}}(x)dx = F_{\mathsf{R}}(s) \tag{A.4}$$

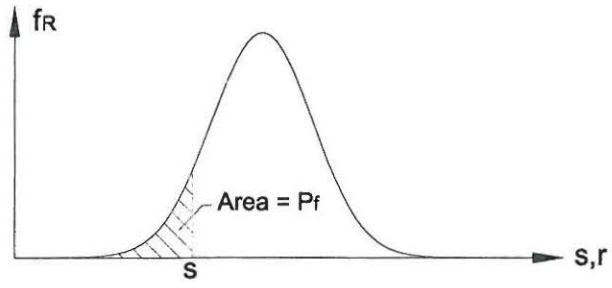


Figure A.2: Density function for the resistance/capacity variable r. Hatched area equals the probability that the "load" exceeds the "resistance".

where

R	Random resistance variable
f_R	Density function of R
F_R	Distribution function of R

If we consider load and resistance to be stochastic variables as well, the probability of failure is given as the product of two independent events, adding up over all possible occurrences (see Fig. A.3 and Eq. A.5).

$$P_f = P(R - S \leq 0) = \sum_{-\infty}^{\infty} F_R(r) \cdot f_S(r)dr = \int_{-\infty}^{\infty} F_R(r) \cdot f_S(r)dr \qquad (A.5)$$

where

$F_R(r)$	probability that R is less or equal to r
$f_S(r)dr$	probability that S lies in the range $r, r + dr$

Eq. A.5 describes the probability of failure in the case of two stochastic variables. In the case of K stochastic variables the probability of failure can be defined as

$$P_f = P(g(\Theta) \leq 0) = \int_{w_f} f_\Theta(\theta)d\theta \qquad (A.6)$$

where f_Θ is the joint density function for the n stochastic variables and w_f is the region of the n-dimensional parameter space where failure occurs. Figure A.4 illustrates the failure surface dividing the parameter region into a safe set and a failure set. The probability of failure is represented by the probability mass in the failure region shown in Figure A.4b.

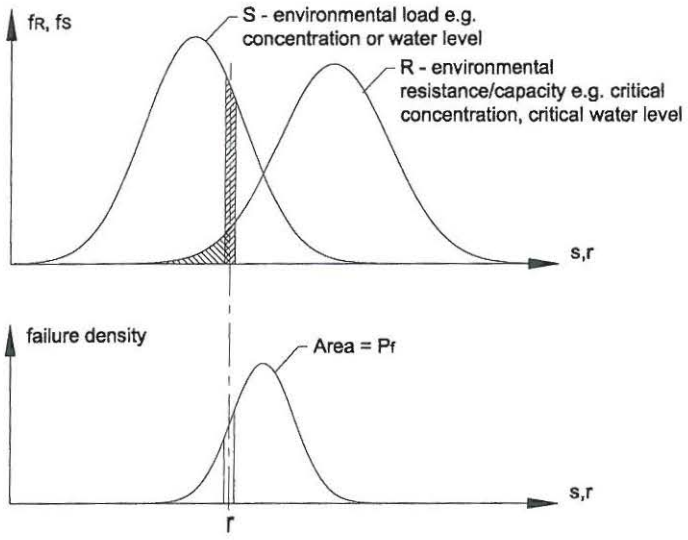


Figure A.3: Probability of failure given the environmental load S and the environmental resistance/capacity R.

Eq. A.6 can be solved analytically only in simple situations. The following sections deal with different methods for approximating $P(g(\mathbf{\Theta}) \leq 0)$.

A.2 The Monte Carlo Method

The Monte Carlo method is one of the simplest and most commonly used reliability methods. The method is unconditionally stable, but tends to be computationally demanding.

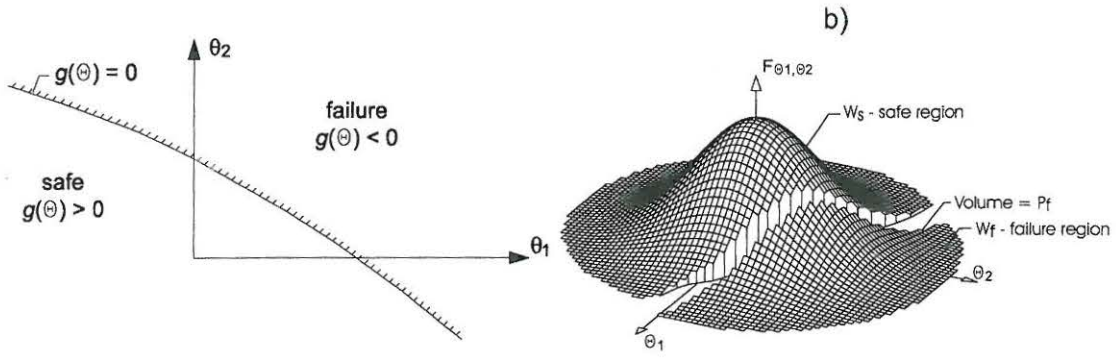


Figure A.4: Illustration of safe and failure regions in a 2-dimensional parameter space - a) failure surface, safe and failure region, b) joint PDF for θ_1 and θ_2 and failure probability

The Monte Carlo method is a simulation method in which a number of realisations N of the K basic variables are randomly generated $\hat{\theta}_1 ... \hat{\theta}_K$ from the stochastic variables $\Theta_1 ... \Theta_K$. For each set of realisations the failure function is evaluated and the number of realisations in the failure region is counted. The probability of failure is simply the relative number of realisations in the failure region and is written as:

$$P\left(g\left(\boldsymbol{\Theta}\right) \leq 0\right) = \frac{1}{N_{sim}} \sum_{n=1}^{N_{sim}} I[g(\hat{\boldsymbol{\theta}}_n)] \tag{A.7}$$

where I is an indicator function defined by

$I[g(\hat{\boldsymbol{\theta}})] = 0$ if $g(\hat{\boldsymbol{\theta}}) > 0$
$I[g(\hat{\boldsymbol{\theta}})] = 1$ if $g(\hat{\boldsymbol{\theta}}) \leq 0$

Example A.1 *Assume that the failure surface in a 2-dimensional parameter space is placed as in Fig. A.5. The probability of failure is to be estimated on the basis of Monte Carlo simulations. 30 realisations of Θ_1 and Θ_2 are randomly generated (dots on Figure A.5).*

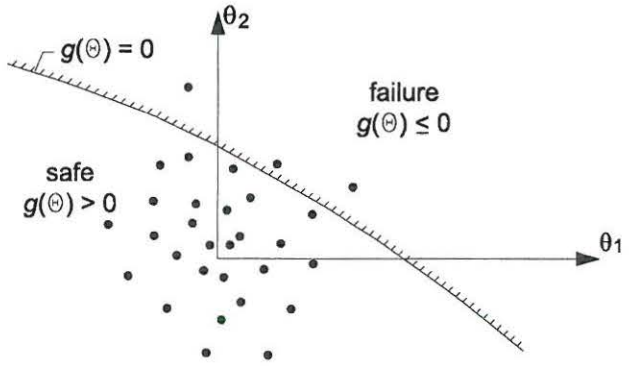


Figure A.5: Graphic representation of the Monte Carlo method

The failure function g is evaluated for each realisation. From Fig. A.5 we can see that 27 realisations lie within the safe area and 3 realisations lie within the failure area; hence the probability of failure is $P_f = \frac{3}{30} = 0.1$

The standard error in a Monte Carlo estimate depends on the level of probability failure level and the number of realisations, N (Rubinstein 1981)

$$s = \sqrt{\frac{P_f(1 - P_f)}{N_{sim}}} \tag{A.8}$$

where P_f is the exact probability of failure, which is unknown, rather than the Monte Carlo estimate of the probability of failure.

A.3 The Latin Hypercube Method

The Latin Hypercube method is a Monte Carlo-like procedure in which the basic idea is to ensure that the total range of each random variable k is sampled. The sample range for each random variable is divided into N_{int} intervals with equal probability (Fig. A.6). A random value within each interval is drawn. This leads to a sample set of N_{int} values for each random variable, k.

$$\hat{\theta} = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1n} & \cdots & \theta_{1N_{int}} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2n} & \cdots & \theta_{2N_{int}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_{k1} & \theta_{k2} & \cdots & \theta_{kn} & \cdots & \theta_{kN_{int}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_{N_{par}1} & \theta_{N_{par}2} & \cdots & \theta_{N_{par}n} & \cdots & \theta_{N_{par}N_{int}} \end{pmatrix} \tag{A.9}$$

The sample set is now combined to produce N_{int} sets of realisations of the basic variables. The N_{int} input vectors are generated from $\hat{\theta}$ by randomly combining values from the intervals of each random variable in such a way that each interval is matched just one time with the interval from each of the other variables (see Fig. A.6).

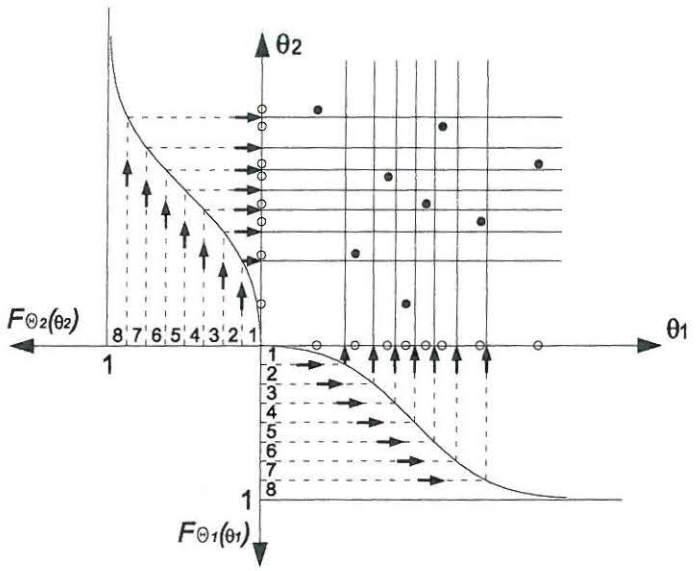


Figure A.6: Latin Hypercube simulation, $N_{int} = 8$, $N_{par} = 2$

This leads to N_{int} combined parameter sets, $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_n, \cdots, \boldsymbol{\theta}_{N_{int}})^T$, and the probability of failure found by evaluating the N_{int} realisations is

$$P\left(g\left(\boldsymbol{\Theta}\right) \leq 0\right) = \frac{1}{N_{int}} \sum_{n=1}^{N_{int}} I[g(\hat{\boldsymbol{\theta}}_n)] \tag{A.10}$$

The procedure can be repeated a number N_{rep} times to improve the quality of the estimate. The estimate of the failure probability is thus

$$P\left(g\left(\boldsymbol{\Theta}\right) \leq 0\right) = \frac{1}{N_{int} N_{rep}} \sum_{m=1}^{N_{rep}} \sum_{n=1}^{N_{int}} I[g(\hat{\boldsymbol{\theta}}_{mn})] \tag{A.11}$$

A.4 First Order Reliability Method

In the First Order Reliability Method the failure function or the limit state function is linearised, and the probability of failure is estimated on the basis of this linearised failure function. The quality of the FORM estimate of the failure probability, unlike the estimate produced by simulation methods, depends only on the linearity/non-linearity of the failure surface. For linear failure surfaces the FORM estimate is exact, whereas for non-linear failure surfaces the given probability of failure is only an estimate.

It is convenient to introduce a safety margin given as $M = g\left(\boldsymbol{\Theta}\right)$. The probability of failure can then be written as

$$P_f = P\left(\mathsf{M} \leq 0\right) = P\left(g\left(\boldsymbol{\Theta}\right) \leq 0\right) = \int_{w_f} f_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} \tag{A.12}$$

A.4.1 Linear Safety Margin

The safety margin can be expressed as a linear function of the basic variables

$$\mathsf{M} = a_0 + a_1 \Theta_1 + \ldots + a_N \Theta_N \tag{A.13}$$

where $a_0, a_1, \ldots, a_N$ are constants. For independent variables the mean and standard deviations of safety margin μ_{M}, σ_{M} are

$$\mu_{\mathsf{M}} = a_0 + a_1 \mu_{\Theta_1} + \ldots + a_N \mu_{\Theta_N} \tag{A.14}$$

$$\sigma_{\mathsf{M}} = \sqrt{a_1^2 \sigma_{\Theta_1}^2 + \ldots + a_N^2 \sigma_{\Theta_N}^2} \tag{A.15}$$

For linear safety margins the reliability index β is introduced as a measure of reliability (Cornell 1966)

$$\beta = \frac{\mu_{\mathsf{M}}}{\sigma_{\mathsf{M}}} \tag{A.16}$$

The safety margin M is normal distributed if the basic variables are normal distributed and the safety margin is linear; thus the probability of failure can be expressed by β

$$P_f = P(\mathsf{M} \leq 0) = P(\mu_{\mathsf{M}} + \mathsf{u}\sigma_{\mathsf{M}} \leq 0) = P\left(\mathsf{u} \leq -\frac{\mu_{\mathsf{M}}}{\sigma_{\mathsf{M}}}\right) = \Phi(-\beta) \tag{A.17}$$

where u is a standard normal distributed variable and Φ is the standard normal distribution function.

Example A.2 *A statistical analysis of the nitrate concentration in a lake was performed and it was found that the nitrate load in the lake, S, could be described using a normal distribution with mean, $\mu_{\mathsf{S}} = 35$ mg/l and standard deviation $\sigma_{\mathsf{S}} = 5$ mg/l. There are plans for a fresh water intake in the lake and the decision-makers want to ensure that there is no health risk in drinking the water. Let us assume that the critical nitrate concentration is uncertain and that it can be described by a normal distribution,* R, *with mean, $\mu_{\mathsf{R}} = 50$ mg/l and standard deviation $\sigma_{\mathsf{R}} = 10$ mg/l. The task now is to determine the health risk involved in consuming lake water. The safety margin for this problem is easily formulated as*

$$\mathsf{M} = g(\mathsf{S}, \mathsf{R}) = \mathsf{R} - \mathsf{S} \tag{A.18}$$

It can be seen that the safety margin is linear and thus that there exists an analytical solution to the problem. The mean and standard deviation of the safety margin is

$$\begin{aligned} \mu_{\mathsf{M}} &= \mu_{\mathsf{R}} - \mu_{\mathsf{S}} = 50 - 35 = 15 \\ \sigma_{\mathsf{M}} &= \sqrt{\sigma_{\mathsf{R}}^2 + \sigma_{\mathsf{S}}^2} = \sqrt{10^2 + 5^2} = 11.2 \end{aligned} \tag{A.19}$$

From Eq. A.16 we can find the reliability index

$$\beta = \frac{\mu_{\mathsf{M}}}{\sigma_{\mathsf{M}}} = \frac{15}{11.2} = 1.34 \tag{A.20}$$

The probability of failure is given as the standard normal distribution in $-\beta$

$$P_f = \Phi(-1.34) = 0.09 \tag{A.21}$$

Thus the health risk is 9 % if the planned fresh water intake is carried out.

A.4.2 Geometrical Interpretation of the Reliability Index

The relationship between the reliability index β and the probability of failure can also be interpreted graphically if we consider two basic independent variables Θ_1 and Θ_2 and a linear failure function

$$g(\boldsymbol{\Theta}) = a_0 + a_1\theta_1 + a_2\theta_2 \tag{A.22}$$

If the normalised stochastic variables u_1 and u_2, with zero expected value and unit standard deviation, are introduced by

$$\mathsf{u}_i = \frac{\boldsymbol{\theta}_i - \mu_{\boldsymbol{\theta}_i}}{\sigma_{\boldsymbol{\theta}_i}}, \qquad i = 1,2 \tag{A.23}$$

then the failure function can be written as

$$\begin{aligned} g(\mathbf{u}) &= a_0 + a_1(\mu_{\Theta_1} + \sigma_{\Theta_1}u_1) + a_2(\mu_{\Theta_2} + \sigma_{\Theta_2}u_2) \\ &= a_0 + a_1\mu_{\Theta_1} + a_2\mu_{\Theta_2} + a_1\sigma_{\Theta_1}u_1 + a_2\sigma_{\Theta_2}u_2 \end{aligned} \tag{A.24}$$

If the reliability index β is introduced, this is equivalent to

$$g(\mathbf{u}) = \beta - \alpha_1 u_1 + \alpha_2 u_2 \tag{A.25}$$

where

$$\beta = \frac{\mu_{\mathsf{M}}}{\sigma_{\mathsf{M}}} = \frac{a_0 + a_1\mu_{\Theta_1} + a_2\mu_{\Theta_2}}{\sqrt{a_1^2\sigma_{\Theta_1}^2 + a_2^2\sigma_{\Theta_2}^2}} \tag{A.26}$$

$$\alpha_i = \frac{-a_i\sigma_{\Theta_i}}{\sqrt{a_1^2\sigma_{\Theta_1}^2 + a_2^2\sigma_{\Theta_2}^2}} \qquad i = 1,2 \tag{A.27}$$

α_i is normalised to satisfy $\sqrt{\alpha_1^2 + \alpha_2^2} = 1$.

Fig. A.7 sketches a linear failure function in a two-dimensional normalised variable **u**-space. The reliability index β is seen as the shortest distance from origin to the failure surface. The $\boldsymbol{\alpha}$ vector points from the origin in the direction of the β-point.

The joint distribution function for u_1 and u_2 is a normalised two-dimensional normal distribution function. It is easily seen that a projection of the probability mass to any vector in the $\mathsf{u}_1 - \mathsf{u}_2$ space will result in a one-dimensional normalised normal distribution. In Fig. A.8 the probability mass of $f_{\mathsf{u}_1\mathsf{u}_2}$ is projected onto the line A-A (Figure A.7). Since the failure surface is linear and perpendicular

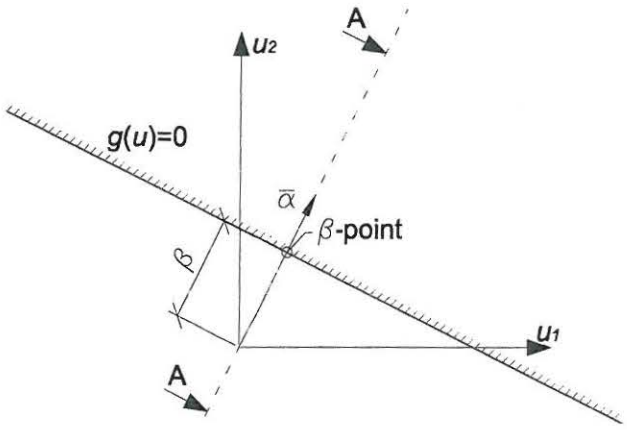


Figure A.7: A two-dimensional normalised variable space with linear failure surface

to the α-vector and the projection line A-A, the probability mass in the failure region must correspond to the probability mass from $\alpha = \beta$ to $\alpha = \infty$ in the normalised one-dimensional normal distribution shown in Fig. A.8, or simply $\Phi(-\beta)$

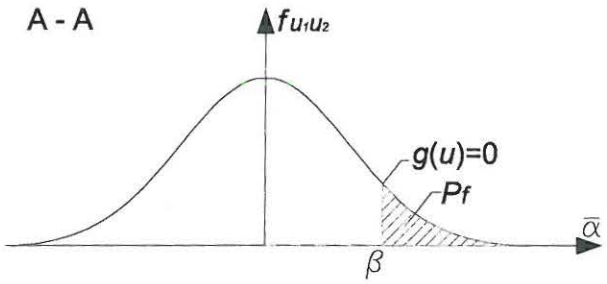


Figure A.8: Probability mass of $f_{u_1u_2}$ projected onto a line with the unit vector α

A.4.3 Non-linear Safety Margin

In most situations the failure surface is non-linear and the safety margin M is thus not normal distributed. An estimate of the reliability index can be obtained by linearising the failure function in the β-point. The estimated probability of failure is then given as $P_f = \Phi(-\beta)$. Figure A.9 shows a non-linear failure function in a two dimensional normalised variable space (N(0,1)) (See section A.4.2 for transformation into normalised variable space.) The failure function is linearised in the β-point.

The error in the approximation of the probability of failure is the probability

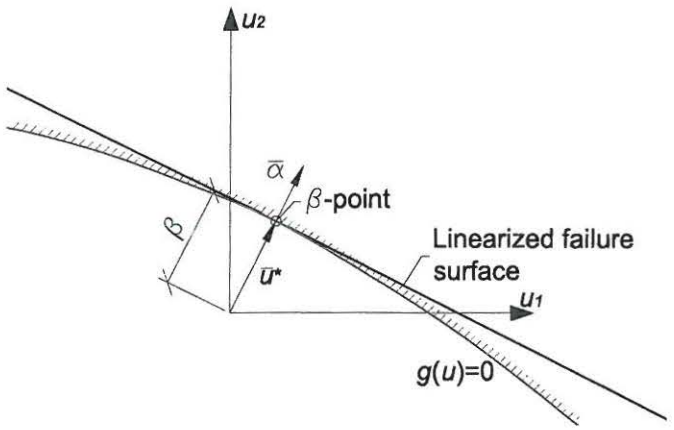


Figure A.9: A two dimensional normalised variable space with non-linear failure surface

mass between the non-linear failure function and the linearised failure function (see Fig. A.9)

The main problem in the FORM method lies in the determination of the β-point. The reliability index is defined by the optimisation problem

$$\beta = \min_{g_u(\mathbf{u})=0} \sqrt{\sum_{i=1}^{N} u_i^2} \tag{A.28}$$

where the solution point for $\mathbf{u}$ is denoted as $\mathbf{u}^*$, see figure A.9

In this text the Steepest Descent method is used to find $\mathbf{u}^*$. At the β-point the following relation must be fulfilled

$$\mathbf{u}^* = \lambda \nabla g_u(\mathbf{u}^*) \tag{A.29}$$

where λ is a proportionality factor. if we assume that a point $\mathbf{u}^0$ close to $\mathbf{u}^*$ is known, i.e.

$$\mathbf{u}^* = \mathbf{u}^0 + \Delta\mathbf{u} \tag{A.30}$$

and use a first order Taylor approximation of $g_u(\mathbf{u})$ in $\mathbf{u}^0$, this gives

$$g_u(\mathbf{u}^*) \simeq g_u(\mathbf{u}^0) + \nabla g_u(\mathbf{u}^0)^T (\mathbf{u}^* - \mathbf{u}^0) = g_u(\mathbf{u}^0) + \nabla g_u(\mathbf{u}^0)^T \Delta\mathbf{u} \tag{A.31}$$

Combining A.29 and A.31 gives

$$g_u(\mathbf{u}^*) \simeq g_u(\mathbf{u}^0) + \nabla g_u(\mathbf{u}^0)^T (\mathbf{u}^* - \mathbf{u}^0) = g_u(\mathbf{u}^0) + \nabla g_u(\mathbf{u}^0)^T (\lambda \nabla g_u(\mathbf{u}^*) - \mathbf{u}^0) \tag{A.32}$$

isolating λ in A.32 results in

$$\lambda = \frac{\nabla g_{\mathbf{u}} \left(\mathbf{u}^0\right)^T \mathbf{u}^0 - g_{\mathbf{u}} \left(\mathbf{u}^0\right)}{\nabla g_{\mathbf{u}} \left(\mathbf{u}^0\right)^T \nabla g_{\mathbf{u}} \left(\mathbf{u}^0\right)} \tag{A.33}$$

The $\mathbf{u}^*$-vector can now be found with the iteration scheme:

1. Make an initial guess concerning $\mathbf{u}^0$

2. Calculate $g_{\mathbf{u}} \left(\mathbf{u}^i\right)$, where i is the current iteration number

3. Calculate $\nabla g_{\mathbf{u}} \left(\mathbf{u}^i\right)$

4. Calculate an improved guess of the β-point using Eq. A.29

$$\mathbf{u}^{i+1} = \lambda \nabla g_{\mathbf{u}} \left(\mathbf{u}^i\right) \tag{A.34}$$

5. Calculate the reliability index corresponding to $\mathbf{u}^{i+1}$

$$\beta^{i+1} = \sqrt{\left(\mathbf{u}^{i+1}\right)^T \mathbf{u}^{i+1}} \tag{A.35}$$

6. Check for convergence of β: if $|\beta^{i+1} - \beta^i| \leq tolerance$ then stop, otherwise proceed to step 2.

A.4.4 Sensitivity

If a unit normal vector $\boldsymbol{\alpha}$ to the failure surface at the β-point $\mathbf{u}^*$ is defined by

$$\boldsymbol{\alpha} = -\frac{\nabla g_{\mathbf{u}}(\mathbf{u}^*)}{|\nabla g_{\mathbf{u}}(\mathbf{u}^*)|} \tag{A.36}$$

then the β-point $\mathbf{u}$ can be written, (see Eq. A.29)

$$\mathbf{u}^* = \beta \boldsymbol{\alpha} \tag{A.37}$$

It should be noted that $\boldsymbol{\alpha}$ is directed toward the failure set. The safety margin corresponding to the tangent hyperplane obtained by linearising the failure function at the β-point can then be written

$$\mathsf{M} = \beta - \boldsymbol{\alpha}^T \mathbf{u} \tag{A.38}$$

Further, by using $\boldsymbol{\alpha}^T\boldsymbol{\alpha} = 1$ it can be seen from Eq. (A.37) that the reliability index β can be written

$$\beta = \boldsymbol{\alpha}^T \mathbf{u}^* = \alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_n u_n \tag{A.39}$$

In the case of fixed α the equation for β can be differentiated

$$\frac{d\beta}{du_i}|_{\mathbf{u}=u^*} = \alpha_i \tag{A.40}$$

i.e. the components of the α-vector can be considered as a measure of the relative importance of the uncertainty in the corresponding stochastic variable on the reliability index. However, it should be noted that for dependent (correlated) basic variables the components in the α-vector cannot be linked to a specific basic variable.

Another important sensitivity measure relating to α_i is the so-called omission sensitivity factor ζ_i suggested by Madsen (1988). This factor gives the relative importance of the reliability index by assuming that the stochastic variable no. i is fixed, i.e. is considered a deterministic quantity. If variable no. i is fixed to the value u_i^0 then the safety margin in the normalised space is written

$$\mathsf{M}_i' = \beta - \alpha_i u_i^0 - \sum_{j=1, j\neq i}^{n} \alpha_j \mathbf{U}_j \tag{A.41}$$

with the reliability index

$$\beta_i' = \frac{\beta - \alpha_i u_i^0}{\sqrt{1-\alpha_i^2}} \tag{A.42}$$

The omission sensitivity factor ζ_i is defined by

$$\zeta_i = \frac{\beta_i'}{\beta} = \frac{1 - \alpha_i u_i^0/\beta}{\sqrt{1-\alpha_i^2}} \tag{A.43}$$

Especially if $u_i^0 = 0$ is chosen then

$$\zeta_i = \frac{1}{\sqrt{1-\alpha_i^2}} \tag{A.44}$$

It can be seen that if $|\alpha_i| < 0.14$ then $\zeta_i - 1 < 0.01$, i.e. the error in the reliability index is less than 1% if a variable with $|\alpha_i| < 0.14$ is fixed.

APPENDIX B

Simulation of random variables

An essential step in simulation methods is the simulation of the outcome of a stochastic variable with arbitrary distribution. Common to the simulation of any type of stochastic variable is the generation of uniform distributed numbers (random numbers). The generation of random numbers is described in section B.1 and in sections B.2 - B.3.1 the generation of the widely-used normal and log-normal distributions is described. Section B.4 and section B.4.1 deals with the generation of correlated variables. A good description of the simulation of a wide range of stochastic variables can be found in Law and Kelton (1991)

B.1 Generation of random numbers

The most commonly used random number generators are not usually real random generators but only pseudo-random generators. The pseudo-random generators use rules for generating a sequence of numbers and the sequence will be repeated after some time. A sequence of random numbers obtained from a pseudo-random generator can always be repeated if the starting conditions are the same.

One of the most widely used methods for simulating pseudo-random numbers is to employ linear congruential generators (LCGs). The pseudo-random numbers are determined sequentially by

$$Z_i = \mathrm{mod}\,(aZ_{i-1} + c, m) \tag{B.1}$$

where m (the modulus), a (the multiplier), c (the increment), and Z_0 (the seed or starting value) are non-negative values. Equation (B.1) says that Z_i is the

remainder of the division of $(aZ_{i-1}+c)$ by m. Thus $0 \leq Z_i \leq m-1$ and to obtain a random number V_i, uniformly distributed from 0 to 1, we let $V_i = Z_i/m$. The sequence of numbers is repeated after at most m steps. The full period m is obtained if (Law and Kelton 1991)

- the only positive integer that exactly divides m and c is 1.
- if q is a prime number (divisible by only itself and 1) that divides m, then q divides a-1.
- if 4 divides m, then 4 divides a-1

On many computers the following generator is used

$$Z_i = \mathrm{mod}\left(89069 Z_{i-1}, 2^{32}\right) \tag{B.2}$$

B.2 Inverse Method

Based on randomly generated numbers, the inverse method simulates any outcome $\hat{\theta}$ of the stochastic variable Θ with distribution $F_\Theta(\theta)$. The inverse method involves two steps:
1. generate an outcome $\hat{v}$ of V
2. determine the outcome of $\hat{\theta}$ as

$$\hat{\theta} = F_\Theta^{-1}\left(F_V\left(\hat{v}\right)\right) = F_\Theta^{-1}\left(\hat{v}\right) \tag{B.3}$$

Figure B.1 illustrates the algorithm graphically. $\hat{v}_1$ and $\hat{v}_2$ are numbers generated from a unit normal distributed variable, U(0,1) and $\hat{\theta}_1$ and $\hat{\theta}_2$ are found from (B.3).

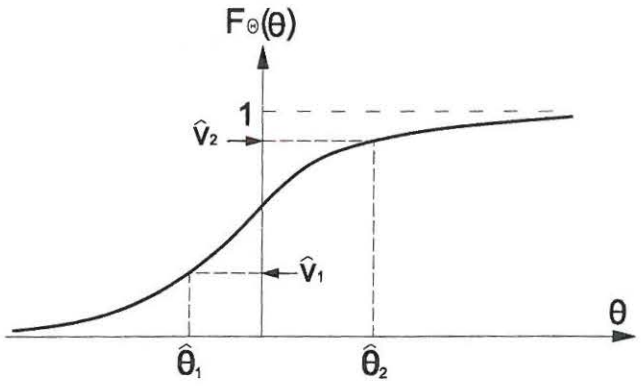


Figure B.1: Illustration of the inverse method

B.3 Box-Muller method for simulating normal distributed numbers

The Box-Muller method generates $\hat{u}_1$ and $\hat{u}_2$ from an independent standardised normal distribution N(0,1)

$$\begin{aligned} \hat{u}_1 &= \sqrt{-2\,ln\,\hat{v}_1}\cos(2\pi\hat{v}_2) \\ \hat{u}_2 &= \sqrt{-2\,ln\,\hat{v}_1}\sin(2\pi\hat{v}_2) \end{aligned} \tag{B.4}$$

where $\hat{v}_1$ and $\hat{v}_2$ are random numbers generated from U(0,1) (e.g. using a linear congruential generator). The transformation from normalised space (u-space) to x-space is given as

$$\hat{\theta} = \mu_\Theta + \hat{u}\cdot\sigma_\Theta \tag{B.5}$$

Example B.1 *Two normal distributed numbers are to be generated from N(5,1) using the Box-Muller method. First we use the linear congruential generator described by Eq.(B.1) with starting value, $z_0 = 123.456.789$ to generate two numbers, v_1, v_2 from the U(0,1) distribution.*

$$\begin{aligned} Z_1 &= mod\left(89069Z\cdot 123456789, 2^{32}\right) = 1056461640 \\ &\Downarrow \\ v_1 &= Z_1/m = 1056461640/2^{32} = 0.2458 \\ \\ Z_2 &= mod\left(89069\cdot 1056461640, 2^{32}\right) = 3838292432 \\ &\Downarrow \\ v_2 &= Z_2/m = 3838292432/2^{32} = 0.8937 \end{aligned} \tag{B.6}$$

Secondly the Box-Muller method is used to generate the two standardised normal distributed numbers , u_1 and u_2.

$$\begin{aligned} \hat{u}_1 &= \sqrt{-2\ln 0.2458}\cos(2\pi\cdot 0.8937) = 1.315 \\ \hat{u}_2 &= \sqrt{-2\ln 0.2458}\sin(2\pi\cdot 0.8937) = -1.036 \end{aligned} \tag{B.7}$$

Finally the normal distributed numbers are found as

$$\begin{aligned} \hat{\theta}_1 &= \mu_\Theta + \hat{u}_1\cdot\sigma_\Theta = 5 + 1.315\cdot 1 = 6.315 \\ \hat{\theta}_2 &= \mu_\Theta + \hat{u}_2\cdot\sigma_\Theta = 5 - 1.036\cdot 1 = 3.964 \end{aligned} \tag{B.8}$$

B.3.1 Simulation of Log-normal distributed numbers

The Log-normal distribution has the property that if $Y \sim \mathrm{N}(\mu,\sigma^2)$ then $e^Y \sim \mathrm{LN}(\mu,\sigma^2)$. Log-normal distributed numbers can be generated from the following two-step algorithm:

- Generate $Y \sim \mathrm{N}(\mu, \sigma^2)$, e.g. by using the Inverse method or Box-Muller method.
- Calculate the log-normal number Θ as $\Theta = e^Y$

μ and σ is not the mean and variance of the $\mathrm{LN}(\mu, \sigma^2)$. If $\Theta \sim \mathrm{LN}(\mu, \sigma^2)$, then the expected value of Θ is $\mu_l = E(\Theta)$ and the variance of Θ is $\sigma_l = Var(\Theta)$. The relation between μ_l and σ_l and μ and σ can be derived using the definition of E[Θ] and VAR[Θ]:

$$\mu = \ln\left(\frac{\mu_l^2}{\sqrt{\sigma_l^2 + \mu_l^2}}\right) \tag{B.9}$$

$$\sigma^2 = \ln\left(\frac{\sigma_l^2 + \mu_l^2}{\mu_l^2}\right) \tag{B.10}$$

B.4 Simulation of correlated variables

Let us assume that the transformation between unit normal distributed numbers, $\mathbf{U}$ and correlated unit normal distributed numbers, $\mathbf{Y}$, with correlation coefficient, $\boldsymbol{\rho}$, can be described as

$$\mathbf{Y} = \mathbf{T}\mathbf{U} \tag{B.11}$$

where $\mathbf{T}$ is a lower triangular transformation matrix.

The covariance matrix $\mathbf{C_Y}$ can be written as

$$\mathbf{C_Y} = E[\mathbf{Y}\ \mathbf{Y^T}] = E[\mathbf{T}\ \mathbf{U}\ \mathbf{U}^T\ \mathbf{T}^T] = \mathbf{T}\ E[\mathbf{U}\ \mathbf{U}^T]\ \mathbf{T}^T = \mathbf{T}\ \mathbf{T}^T = \boldsymbol{\rho} \tag{B.12}$$

and the elements of the transformation matrix $\mathbf{T}$ can be found from $\mathbf{T}\ \mathbf{T}^T = \boldsymbol{\rho}$ as

$$\begin{aligned} T_{i,j} &= \rho_{i,j} && \text{if } i = 1 \\ T_{i,j} &= \sqrt{1 - \textstyle\sum_{k=1}^{i-1} T_{k,j}^2} && \text{if } i \neq 1 \wedge i = j \\ T_{i,j} &= \rho_{i,j} - \frac{\sum_{k=1}^{i-1} T_{k,i} T_{k,j}}{T_{i,i}} && \text{if } i < j \end{aligned} \tag{B.13}$$

The transformation from uncorrelated, unit normal distributed variables, $\mathbf{U}$ to normal distributed variables, $\boldsymbol{\Theta}$ can be written as

$$\boldsymbol{\Theta} = \boldsymbol{\mu}_{\boldsymbol{\Theta}} + \mathbf{DTU} \tag{B.14}$$

where $\mathbf{D}$ is a diagonal matrix with standard deviations of $\boldsymbol{\Theta}$ in the diagonal.

B.4.1 Simulation of stochastic fields

Let us assume that the spatial correlation can be described by an exponential decaying correlation structure. The spatial correlation between the numerical cell p_1 and numerical cell p_2 is then given as

$$\rho_{p_i,p_j} = \exp\left(\frac{-\sqrt{(x_{p_1} - x_{p_2})^2 + (y_{p_1} - y_{p_2})^2}}{I}\right) \tag{B.15}$$

where $x_{p_1}, x_{p_2}, y_{p_1}$ and y_{p_2} are the coordinates of the centre of cell p_1 and p_2. I is the integral scale.

The correlation matrix, $\boldsymbol{\rho}$, describing the correlation between all numerical cells is constructed, and the procedure presented in section B.4 is used to calculate the realisation of the correlated stochastic variable in each cell.

References

Abbott, M., J. Bathurst, J. Cunge, P. O'Connel, and J. Rasmussen (1986). An introduction to the european hydrological system - systeme hydrologique europeen, "SHE", 2: Structure of a Physically-Based, distributed modelling system. *Journal of Hydrology 87*, 61–77.

Bard, Y. (1974). *Nonlinear Parameter Estimation.* Academic Press, Inc.

Bear, J. (1972). *Dynamics of Fluids in Porous Media.* American Elsevier Publishing Company, Inc.

Bear, J. (1979). *Hydraulics of Groundwater.* McGraw-Hill Inc.

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in water resources 16*, 41–51.

Beven, K. and A. Binley (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes 6*, 279–298.

Beven, K., A. Calver, and E. Morris (1987). The institute of hydrology distributed model. *Institute of Hydrology Report 98, Wallingfort, UK.*

Bjarnov, S. (1987). Beregning af medianminimum på grundlag af årsminimumsserier. Hedeselskabets Forskningsvirksomhed, Hedeselskabet, Beregning nr. 32.

Blicher, A. S. (1991). Usikkerhed på bearbejdning af data fra vandløbsstationer (revideret udgave) - publikation nr. 1. Fagcenter for hydrometriske data, Hedeselskabet.

Carrera, J. and S. P. Neuman (1986a, February). Estimation of aquifer parameters under transient and steady state conditions: 1. maximum likelihood method incorporating prior information. *Water Resourses Research 22*(2), 199–210.

Carrera, J. and S. P. Neuman (1986b, February). Estimation of aquifer parameters under transient and steady state conditions: 2. uniqueness, stability, and solution algorithms. *Water Resourses Research 22*(2), 211–227.

Christensen, S. and R. L. Cooley (1999). Evaluation of prediction intervals for expressing uncertainties in groundwater flow models predictions. *Water Resources Research 35*(9), 2627–2639.

Christensen, S., K. Rasmussen, and K. Møller (1998). Prediction of regional ground water flow to streams. *Ground Water 36*(2), 351–360.

Cooley, R. L. (1977). A method of estimating parameters and assessing reliability for models of steady state groundwater flow - 1. theory and numerical properties. *Water Resources Research 13*(2), 318–324.

Cooley, R. L. (1979). A method of estimating parameters and assessing reliability for models of steady state groundwater flow - 2. application of statistical analysis. *Water Resources Research 15*(3), 603–617.

Cornell, C. A. (1966). A probabillity-based structural code. *ACI 66*, 974–985.

Coulibaly, P., F. Anctil, R. Aravena, and B. Bobee (2001). Artificial neural network modeling of water depth fluctuations. *Water Resources Research 37*(4), 885–896.

Danmarks Statistik (2002). Forbruget af drikkevand 2000, statistiske efterretninger (in danish). Technical report, Danmarks Statistik.

Duan, Q., S. Sorooshian, and V. Gupta (1992). Effective and efficient global optimization for conceptual Rainfall-Runoff models. *Water Resources Research 28*(4), 1015–1031.

Dubois, D. and H. Prade (1980). *Fuzzy Sets and Systems: Theory and Applications*. Academic Press.

Feyen, L., K. Beven, F. Smedt, and J. Freer (2001, March). Stochastic capture zone delineation within the generalized likelihood uncertainty estimation methodology: Conditioning on head observations. *Water Resources Research 37*(3), 625–638.

Feyen, L., P. R. JR, J. Gómez-Hernández, K. Beven, and F. D. Smedt (2002). The worth of transmissivity and head data in the prediction of well capture zones. In K. Kovar and Z.Hrkal (Eds.), *Preceedings of the 4th International Conference on Calibration and Reliability in Groundwater Modelling, ModelCare 2002*, pp. 477 – 480. ACTA Universitatis Carolinae - Geologica.

Freeze, R. A. (1975). A stochastic-conceptional analysis of one-dimentional groundwater flow in non-uniform porous media. *Water Ressources Research 1*(11), 724–741.

Gelhar, L. W. (1986). Stochastic subsurface hydrology - from theory to applications. *Water Resources Research 22*(9), 135S–145S.

Harbaugh, A., E. Banta, M. Hill, and M. McDonald (2000). MODFLOW-2000, the U.S. geological survey modular Ground-Water model - user guide to modularization concepts and the Ground-Water flow process. Technical Report Open-File Report 00-92, U.S. Geological Survey.

Hill, M. C. (1992). A computer program (MODFLOWP) for estimating parameters of a transient, Three-Dimensional, Ground-Water flow model us-

ing nonlinear regression. Technical Report Open-File Report 91-484, U.S. Geological Survey.

Hill, M. C. (1994). Five computer programs for testing weighted residuals and calculating linear confidence and prediction invervals on results from the Ground-Water Parameter-Estimation computer program MODFLOWP. Technical Report Open-File Report 93-481, U.S. Geological Survey.

Jensen, J. B. and K. Schaarup-Jensen (2002). Delineation of capture zones by an integrated Surface/Subsurface model using the GLUE methodology. In *Calibration and Reliability in Gruondwater Modelling: A Few Steps Closer to Reality (Proceedings of ModelCare2002, Prague, Czech Republic, 17-20 June 2002)*. IAHS Publ. no. 277.

Koskinen, L., M. Laitinen, J. Lofman, K. Meling, and F. Meszaros (1996). FEFLOW: a finite element code for simulating groundwater flow, heat transfer and solute transport. In P. Zannetti and C. Brebbia (Eds.), *Development and Application of Computer Techniques to Environmental Studies VI.*

Law, A. M. and W. D. Kelton (1991). *Simulation Modeling and Analysis.* McGraw-Hill.

Leeuwen, M. V. (2000). *Stochastic Determination of Well Capture Zones Conditioned on Transmissivity Data.* Ph. D. thesis, University of London.

Madsen, H. (1988). Omission sensitivity factors. *Structual Safety 5*(1), 33–45.

Madsen, H. and M. Kristensen (2002). A Multi-Objective calibration framework for parameter estimation in the MIKE SHE integrated hydrological modelling system. In K. Kovar and Z. Hrkal (Eds.), *Calibration and Reliability in Gruondwater Modelling: A Few Steps Closer to Reality (Proceedings of ModelCare2002, Prague, Czech Republic, 17-20 June 2002)*, pp. 270–273.

McDonald, M. and A. Harbaugh (1988). *A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model.* U.S. Geological Survey. manual 83-875.

Mehl, S. and M. Hill (2002). Development and evaluation of local grid refinement method for Block-Centered Finite-Difference groundwater models using shared nodes. *Advances in Water Ressources 25*, 497–511.

Melching, C. S. (1995). Reliability estimation. In V. P. Singh (Ed.), *Computer Models of Watershed Hydrology*, pp. 69 – 118. Water Resources Publication.

Miljøstyrelsen (2000). Vandmiljø 2000 - status og perspektiver for et renere vandmiljø, redegørelses nr. 7. (in danish). Technical report, Miljøstyrelsen.

Neuman, S. (2002). Accounting for conceptual model uncertainty via maximum likelihood bayesian averaging. In K. Kovar and Z. Hrkal (Eds.), *Preceedings of the 4th International Conference on Calibration and Reliability in Groundwater Modelling - ModelCARE 2002*, pp. 529–534. ACTA Universitatis Carolinae - Geologica 2002.

Nielsen, S. and E. Hansen (1973). Numerical simulation of the Rainfall-Runoff process on daily basis. *Nordic Hydrology 4*, 171–190.

Oppe, T. C., W. D. Joubert, and D. R. Kincaid (1988). *NSPCG User's Guide - Version 1.0.* Center of Numerical Analysis, The University of Texas at Austin. http://www.ma.utexas.edu/CNA/NSPCG/.

Prickett, T., T. Naymik, and C. G. Lonnquist (1981). *A "Random-Walk" Solute Transport Model for Selected Groundwater Quality Evaluations.* Illinois State Water Survey. Bulletin 65.

Refsgaard, J. C. (1996). *Distributed Hydrological Modelling*, Chapter 2 - Terminology, Modelling Protocol and Classification of Hydrological Model Codes, pp. 17–39. Kluwer Academic Publishers.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method.* John Wiley and Sons.

Ruppert, J. (1995). A delaunay refinement algorithm for quality 2-dimensional mesh generation. *Journal of Algorithms 18*(3), 548–585.

Seber, G. and C. Wild (1989). *Nonlinear Regression.* John Wiley & Sons, Inc.

Sitar, N., J. Cawlfield, and A. Kiureghian (1987). First-Order reliability approach to stochastic analysis of subsurface flow and contaminant transport. *Water Ressources Research 23*(5), 794–804.

Sonnenborg, T. (2001). Kalibrering af strømningsmodel. In H. Henriksen, J. Refsgaard, T. Sonnenborg, P. Gravesen, A. Brun, A. Refsgaard, and K. Jensen (Eds.), *Ståbi I Grundvandsmodellering*, Chapter 10. GEUS (in danish).

Vanderkwaak, J. and E. Sudicky (2000). Application of a Physically-Based numerical model of surface and subsurface water flow and solute transport. *Calibration and reliability groundwater modelling. IAHS publication no. 265.*, 515–523.

Wu, M. and J. Cawlfield (1992). Probabilistic sensitivity and modeling of Two-Dimensional transport in porous media. *Stochastic Hydrology and Hydraulics 6*, 103–121.

Yapo, P., H. V. Gupta, and S. Sorooshian (1998). Multi-Objective global optimization for hydrologic models. *Journal of Hydrology 204*, 83–97.

Yeh, W. W.-G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research 22*(2), 95–108.

Zimmermann, H.-J. (1991). *Fuzzy Set Theory and its Applications.* Kluwer Academic Publisher.

PARAMETER AND UNCERTAINTY ESTIMATION IN GROUNDWATER MODELLING

Jacob Birk Jensen

Thesis accepted for the degree of Doctor of Philosophy at the Faculty of Engineering and Science, Aalborg University, Denmark. Publicity defended at Aalborg University, January 10, 2003.

The data basis on which groundwater models are constructed is in general very incomplete. The weak and ambiguous data basis may lead to several conceptualizations and parameterizations that all are acceptable in reproducing the system under consideration. Facing this problem we meet the limitations in the classical regression approach where uniqueness regarding model and parameters are assumed.

An alternative approach - the Generalized Likelihood Uncertainty Estimation (GLUE) Methodology - has earlier been suggested in order to account for limitations in the classical regression approach.

In this thesis GLUE is adapted for estimation of the uncertainty related to capture zone modelling on a synthetic setup. Furthermore, the methodology is applied to a regional aquifer system, where head and stream flow predictions are compared with results achieved from the classical regression approach.

Hydraulics & Coastal Engineering Laboratory
Department of Civil Engineering
AALBORG UNIVERSITY
ISSN 0909-4296
Series Paper No. 23